AD 404 580

63-3-4

QUARTERLY PROGRESS REPORT

No. 69

APRIL 15, 1963

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory in which faculty members and graduate students from numerous academic departments conduct research.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the U.S. Army, the U.S. Navy (Office of Naval Research), and the U.S. Air Force (Office of Scientific Research), under Contract DA36-039-sc-78108, Department of the Army Task 3-99-25-001-08; and in part by Grant No. DA-SIG-36-039-61-G14.

Partial support for work in Plasma Dynamics is provided by the U.S. Atomic Energy Commission under Contract AT(30-1)-1842; the National Science Foundation (Grant G-24073); the U.S. Air Force (Electronic Systems Division) under Contract AF 19(604)-5992; and the U.S. Air Force (Aeronautical Systems Division) under Contract AF 33(616)-7624.

Partial support for work in Communication Sciences is provided by the National Science Foundation (Grant G-16526) and the National Institutes of Health (Grant MH-04737-02).

Additional support of specific projects is acknowledged in footnotes to the appropriate sections.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

QUARTERLY PROGRESS REPORT No. 69

April 15, 1963

Submitted by: H. J. Zimmermann
G. G. Harvey
W. B. Davenport, Jr.

TABLE OF CONTENTS

CONTENTS

CONTENTS

PERSONNEL

Administration

Professor H. J. Zimmermann, Director
Professor G. G. Harvey, Associate Director
Professor W. B. Davenport, Jr., Associate Director
Mr. R. A. Sayers, Personnel and Business Manager

Advisory Committee

Dean G. S. Brown (Ex officio)
Professor W. W. Buechner
Professor W. B. Davenport, Jr.
Professor P. Elias
Professor G. G. Harvey
Professor A. G. Hill
Professor I. W. Sizer
Professor H. J. Zimmermann (Chairman)

Faculty

Allis, W. P. (Absent)
Arden, D. N. (Absent)
Barrett, A. H.
Bekefi, G.
Bers, A.
Billman, K. W.
Bitter, F.
Bose, A. G.
Brazier, Mary A. B. (Visiting)
Brown, S. C.
Carabateas, E. N.
Chomsky, A. N.
Chorover, S. L.
Chu, L. J.
Cochran, J. F.
Dennis, J. B.
Dupree, T. H.
Eden, M.
Edgerton, H. E.
Elias, P.
Fano, R. M.
Freedman, S. I.
Gallager, R. G.
Garland, C. W.
Getty, W. D.
Goldstein, M. H., Jr.
Graham, J. W.
Green, D. M.
Guillemin, E. A.
Gyftopoulos, E. P.
Halle, M.
Hammes, G. G.
Harvey, G. G.
Hatsopoulos, G. N. (Absent)
Haus, H. A.
Heinz, J. M.
Heiser, W. H.
Hennie, F. C., III
Hill, A. G. (Absent)
Hoffman, M. A.
Hofstetter, E. M.
Huffman, D. A.
Ingard, K. U.
Jackson, W. D.
Jacobs, I. M.
Jakobson, R.
Javan, A.
Kerrebrock, J. L.
King, J. G.
Kinsey, J. L.
Klima, E. S.
Kyhl, R. L.
Lee, Y. W.
Lettvin, J. Y.
Lidsky, L. M.
Manders, A. M.
Mason, S. J.
Matthews, G. H.
Maxwell, E. (Visiting)
Meissner, H. P.
Melcher, J. R.
Melzack, R.
Minsky, M. L.
More, T., Jr.
Nottingham, W. B.
Oates, G. C.
Peake, W. T.
Penfield, P. L., Jr.
Penhune, J. P.
Perry, C. H.
Putnam, H.
Rafuse, R. P.
Reiffen, B. (Visiting)
Reynolds, J. M.
Rose, D. J.
Rosenblith, W. A.
Sakrison, D. J.
Schetzen, M.
Schreiber, W. F.
Schwab, W. C.
Searle, C. L.
Shannon, C. E.
Shapiro, A. H.
Siebert, W. McC.
Smith, J. L., Jr.
Smullin, L. D.
Stevens, K. N. (Absent)
Stickney, R. E.
Stockham, T. G., Jr.
Strandberg, M. W. P.
Swets, J. A. (Absent)
Teager, H. M.
Teuber, H. L.
Thornton, R. D.
Tisza, L. (Absent)
Townes, C. H.
Troxel, D. E.
Van Trees, H. L., Jr.
Wall, P. D.
Warren, B. E.
Waugh, J. S.
Whitehouse, D. R.
Whitney, W. M.
Woodson, H. H.
Wozencraft, J. M.
Young, L. R.
Zacharias, J. R.
Zimmermann, H. J.

PERSONNEL

Instructors

Alter, R.
Bohacek, P. K.
Bruce, J. D.
Cooper, R. S.
Crystal, T. H. (Absent)
Dean, L. W., III
East, D. A.
Hall, J. L., II
Kellner, W. G.
Kennedy, R. S.
Kincaid, T. G.
Kliman, G. B.
Koskinen, M. F.
Krishnayya, J. G.
Liu, C. L.
MacDonald, J. S.
Oppenheim, A. V.
Parente, R. B.
Perlmutter, D. M.
Pierson, E. S.
Schneider, A. J.
Shavit, A.

Lecturers

Pitts, W. H.
Rines, R. H.
Rose, G. A. (Visiting)
Stroke, G. W.
Stroke, H. H.

Research Associates

Barlow, J. S.
Bishop, P. O.
Bradley, L. C., III
Burke, J. J., Jr.
Cavaggioni, A.
Cerrillo, M. V.
Dupress, J.
Fiocco, G.
Gerstein, G. L.
Gesteland, R. C.
Hermann, H. T.
House, A. S.
Kilmer, W. L.
Stark, L.
Thompson, E.
Yngve, V. H.

Staff Members

Anderson, A. T., III
Arndt, R. W.
Badessa, R. S.
Bates, V. J.
Blinn, J. C., III
Bosche, Carol M.
Brown, R. M.
Campbell, Elizabeth J.
Charney, Elinor K.
Clark, W. A., Jr. (Lincoln Laboratory)
Clayton, R. J.
Congleton, Ann
Crist, A. H.
Darlington, J. L.
Duffy, D. F.
Edwards, D. J.
Farley, B. G. (Lincoln Laboratory)
Ferretti, E. F.
Garber, M.
Giberman, E.
Goodall, M. C.
Greibach, Rita J.
Hewitt, J. H.
Howland, B. H. (Lincoln Laboratory)
Ingersoll, J. G.
Ingraham, E. C.
Kannel, Muriel
Katz, J. J.
Kerllenevich, N.
Keyes, R. V., Jr.
Kiang, N. Y-S.
Kuiper, J. W.
Kusse, B. R.
Lindsley, V. Susan
Majumdar, S. K.
McCarthy, J. J.
McCulloch, W. S.
Mulligan, W. J.
O'Brien, F. J.
O'Rourke, Ann M.
Pennell, Martha M.
Percival, W. K.
Pitts, W. H.
Postal, P. M.
Rao, B. D. N.
Rosebury, F.
Ryan, L. W.
Saltalamacchia, A. J.
Sandel, T. T. (Lincoln Laboratory)
Satterthwait, A. C.
Sayers, R. A.
Shiffman, C. A.
Smith, P. L.
Stroke, H. H.
Svihula, G. F.
Szoke, A.
Thomas, Helen L.
Vidale, Eda B.
Viertel, J. J.
Whitney, R. E.
Willis, P. A.
Yaqub, M.

PERSONNEL

Visitors

Baker, F. H.
(Postdoctoral Fellow of the National Institutes of Health)
Fields, H.
Gross, C.
(Postdoctoral Fellow of the National Institutes of Health)
Hall, R. D.
(Postdoctoral Fellow of the National Institute of Mental Health)
Imai, N.
Isami, Y. (Fulbright Research Scholar in Linguistics)
Keyser, S. J.
Kornacker, K.
(Postdoctoral Fellow of the National Institutes of Health)
Menyuk, Paula
(Postdoctoral Fellow of the National Institutes of Health)
Molnar, C. E.
Parks, J. H.
Rojas, J. A.
(Postdoctoral Fellow of the National Institutes of Health)
Rosenbloom, D.
Schiller, P. H.
(Postdoctoral Fellow of the National Institutes of Health)
Sellin, K. G.
(UNESCO Fellow)
Siatkowski, Z.
Toschi, R. (Fulbright Research Scholar)
Troelstra, A.
(NATO Science Fellow)
Walker, D. E.
Walter, W. T.
Wu, K. (R.L.E. Postdoctoral Fellow)
Yasui, S.

Graduate Students

Adcock, T. G.
Addison, R. C., Jr.
(Raytheon Fellow)
Aldrich, J. A.
(RCA Fellow)
Algazi, V. R.
Allen, R. J.
Alvarez de Toledo, F.
Anderson, T. M.
(USN Fellow)
Andrews, J. M., Jr.
Andrews, M. L.
Arbib, M. A.
Arnstein, D. S.
Arunasalam, V.
Austin, M. E.
(National Science Foundation Fellow)
Axelrod, F. S.
Badrawi, M. T.
Bartsch, R. R.
Bauer, R. F.
Berlekamp, E. R.
(National Science Foundation Fellow)
Bever, T. G.
Black, W. L.
(Teaching Assistant)
Blum, M.
Bobrow, D. G.
Botha, D. G.
Brabson, B. B.
(National Science Foundation Fellow)
Brassert, W. L.
Brendel, P. J.
(National Science Foundation Fellow)
Breon, R. K.
(Teaching Assistant)
Briggs, R. J.
(RCA Fellow)
Brown, J. E. (Porter Fellow of the American Physiological Society)
Buntschuh, C. D.
Burgiel, J. C.
Burns, S. K.
(Teaching Assistant)
Bush, A. M.
Cahlander, D. A.
Capranica, R. R.
(Communications Development Training Program Fellow of Bell Telephone Laboratories, Inc.)
Chamberlain, S. G.
Chandra, A. N.
Chapman, J. C.
Cheadle, G.
(USAF Fellow)
Chiao, R. Y.
Clarke, J. F.
Clemens, J. K.
(Teaching Assistant)
Coccoli, J. D.
Cogdell, J. R.
Coggins, J. L.
Cornew, R. W.
Corti, E.
Crowley, J. M.
(U.S. AEC Fellow)
Cunningham, J. E.
Davis, J. A.
(National Science Foundation Fellow)
Dethlefsen, R.
Diamond, B. L.
Domen, J. K.
Drouet, M. G. A.
Dym, H.
Ebert, P. M.
(National Science Foundation Fellow)
Ellis, J. R., Jr.
Falconer, D. D.
(Whitney Fellow)
Fetz, E. E.
Flannery, D. L.
(General Motors Fellow)
Fohl, T.
Forney, G. D., Jr.
Fraim, F. W., IV
Fraser, J. B.
Gann, A. G.
(Raytheon Fellow)
Garmire, Elsa M.
Garosi, G. A.
Gerry, E. T.
Gerstmann, J.
(National Science Foundation Fellow)
Ghosh, S. K.

Graduate Students (continued)

Glenn, W. H., Jr.
(Hughes Aircraft Fellow)
Golub, R.
(I. T. T. Fellow)
Goodman, L. M.
(Sloan Fellow)
Gothard, N.
Gray, C. E.
(Teaching Assistant)
Gray, P. R.
(Teaching Assistant)
Grayzel, A. I.
Greata, J. M.
(Teaching Assistant)
Greenspan, R. L.
Gronemann, U. F.
Guttrich, G. L.
Hall, Barbara C.
(National Science Foundation Fellow)
Halverson, W. D.
Hambrecht, F. T.
Hart, T. P.
Hartenbaum, B. A.
Hegblom, E. R.
Henke, W. L.
(National Science Foundation Fellow)
Heywood, J. B.
Holsinger, J. L.
Hooper, E. B., Jr.
Houk, J. C., Jr.
• (IBM Fellow)
Hsi, C-F. G.
Hsieh, H. Y.
Huang, T. S.
(Teaching Assistant)
Huibonhoa, R.
(Teaching Assistant)
Ingraham, J. C.
(National Science Foundation Fellow)
Ivie, E. L. (National Science Foundation Fellow)
Jameson, P. W.
(Schlumberger Fellow)
Johnston, W. D., Jr.
Jones, J. S.
Kachen, G. I., Jr.
(Teaching Assistant)
Kamen, E. L.
Karvellas, P. H.
Katona, P. G.
Kellen, P. F.
Ketterer, F. D.
Kiparsky, R. P. V.
Kniazzeh, A. G. F.
Kocher, D. G.
Kramer, A. J.
(Teaching Assistant)
Kronquist, R. L.
Kukolich, S. G.
Kuroda, S. Y.
(Whitney Fellow)
Langendoen, D. T.
Lastovka, J. B.
Lee, K. S.
Lercari, R. F.
Lerman, S. H.
Levine, J. D.
Levine, R. C.
(Teaching Assistant)
Levison, W. H.
Lewis, A. T.
Liégeois, F. A.
Lightner, T. M.
Little, R. G.
Llewellyn-Jones, D. T.
Lontai, L. N.
Lowry, E. S.
Luce, D. A.
Luckham, D. C.
Luther, A. H.
Lutz, M. A.
Macomber, J. D.
Mandics, P. A.
Mark, R. G.
Masek, T. D.
Mayo, J. W.
McCawley, J. D.
McMorris, J. A., II
Mendell, L.
Mendelson, J.
Mermelstein, P.
Morse, D. L.
Mussi, R. N. F.
Naves, F.
Nelson, D. E.
(Teaching Assistant)
Nolan, J. J., Jr.
Nowak, R. T.
Oder, R. R.
Olsen, J. H.
Omura, J. K.
(National Science Foundation Fellow)
Paul, A. P.
Pauwels, H. (Belgium-American Educational Foundation Fellow)
Petrie, L. M., Jr.
(Teaching Assistant)
Pfaff, D. W.
Pfeiffer, R. R.
Phipps, C. R., Jr.
Pilla, M. A.
(Teaching Assistant)
Prabhu, V. K.
Praddaude, H. C.
Pruslin, D. H.
Puri, S.
(Teaching Assistant)
Qualls, C. B.
Radoski, H. R.
Rahmani, Mahin
Redi, O.
Reid, M. H.
Reisman, S. S.
(National Science Foundation Fellow)
Richters, J. S.
(National Science Foundation Fellow)
Rizzo, J.
Roberts, L. G.
Rogoff, G. L.
Rook, C. W., Jr.
Rosenbaum, P. S.
Rosenshein, J. S.
Rowe, A. W.
Rummler, W. D.
Sachs, M. B.
(Sloan Fellow)
Sandberg, A. A.
Savage, J. E.
Schlossburg, H.
Schuler, C. J., Jr.
Serafim, P. E.
Shaffner, R. O.
(Teaching Assistant)
Shane, J. R.
(Kennicot-Copper Fellow)
Shimony, U.
Simon, R. W.
Simpson, J. I.
Sklar, J. R.
Smith, W. S., Jr.
Smith, W. W.
Smythe, D. L., Jr.
(Teaching Assistant)
Snodderly, D. M., Jr.
Sobel, I.
Solbes, A.
Spangler, P. S.

Graduate Students (continued)

Staelin, D. H.
Stanten, S. F. (Hughes Aircraft Fellow)
Steele, D. W. (Sloan Fellow)
Steinbrecher, D. H.
Stiglitz, I. G.
Strong, W. J.
Sutherland, W. R. (National Science Foundation Fellow)
Taub, A. (National Institutes of Health Fellow)
Taylor, T. E. (I.C.T. Fellow)
Tepley, N.
Theodoridis, G.
Thiele, A. A.
Thomae, I. H. (National Science Foundation Fellow)
Thomas, L. C. (National Science Foundation Fellow)
Thornburg, C. O., Jr.
Tobey, M. C., Jr. (Teaching Assistant)
Tomlinson, W. J., III
Topaz, J. M.
Tretiak, O. J.
Tse, F. Y-F.
Tulenko, J. S. (U.S. AEC Fellow)
Ulrich, P. B.
Vanderweil, R. G.
Van Horn, E. C., Jr. (Teaching Assistant)
Wade, C. G.
Wagner, C. F. (Teaching Assistant)
Waletzko, J. A.
Wang, G. Y-C.
Ward, R. L. (IBM Fellow)
Wasserman, G. S.
Weiss, T. F.
Welch, J. R. (Teaching Assistant)
Whitman, E. C.
Wickelgren, G. L.
Wiederhold, M. L.
Wilde, D. U. (Teaching Assistant)
Wilde, G. R.
Wilensky, S.
Williamson, R. C.
Willke, H. L., Jr.
Wilson, G. L. (Teaching Assistant)
Wilson, W. J. (Teaching Assistant)
Winett, J. M.
Wissmiller, J. C. (National Science Foundation Fellow)
Witting, H. L. (U.S. AEC Fellow)
Witting, J. M. (I.T.T. Fellow)
Witzke, K. G.
Wolf, R. P.
Woo, J. C.
Wright, B. L.
Wright, J. S.
Yarnell, C. F.
Yudkin, H. L.
Zeiger, H. P.
Zuber, B. L.
Zuckerman, B. M.
Zwicky, A. M., Jr. (National Defense Education Act Fellow)

Undergraduates

Aaron, A. M.
Aponick, A. A.
Bayles, R. U.
Brincko, A. J.
Byers, F. H.
Colodny, Susan F.
Deininger, C. C.
Diephuis, R. J.
Ecklein, D.
Frasco, L. A.
Gadzuk, J. W.
Grabowski, R. E.
Guinan, J. J., Jr.
Hamlin, R. M., Jr.
Hirschfeld, R. A.
Johnson, I. S. C.
Koons, H. C.
Libman, S. M.
Lindes, P.
Mann, C. A., Jr.
Marchese, P. S., Jr.
Murray, M. J.
Nelson, G. P.
Ng, L. C.
Norris, R. B., Jr.
O'Halloran, W. F., Jr.
Parchesky, J.
Perrolle, P. M.
Peterson, B. A.
Piner, S. D.
Rome, J. A.
Rush, R. D.
Scott, T. A.
Siemens, P. J.
Smith, D. F.
Speer, E. R., Jr.
Stark, M. F.
Wan, A. C. M.
Weidner, M. Y.
Williams, J. A.
Wolfberg, M. S.
Yap, B. K.
Zucker, R. S.

Senior Thesis Students

Addis, J. L.
Arnold, P. G.
Baecker, R. M.
Baugh, W. H., III
Bender, M. H.
Books, G. E.
Butterfield, B. O.
Carson, J. F.
Chu, Y. H.
Cohn, T. E.
Cooper, W. W., IV
DiGregorio, R. F.
Dralle, A. V.
Einolf, C. W., Jr.
Ellms, S. R.

PERSONNEL

Senior Thesis Students (continued)

Emerson, S. T.
Epstein, M. R.
Evans, S. A.
Feezor, M. D.
Fevrier, A.
Fiory, A. T.
Flicker, J. K.
Graham, J. D.
Hadden, W. J., Jr.
Hassan, A. R.
Hegyi, D. J.
Hiller, D. E.
Hoffman, E. J.
Hufford, J. H.
Jansen, R. H.
Jordan, H. D.
Joseph, D. K.
Kinzer, T. J., III
Kosdon, S. J.
Kotler, S. J.
Landowne, D.
Langbein, D.
Lennon, W. J.
Lynch, J. T.
McClees, H. C., Jr.
MacDonald, J. F. P.
Manheimer, W. M.
Marble, C. W.
Marmon, W. C.
Meyer, J. S.
Miller, S. W.
Milner, P.
Mudama, E. L.
Musslewhite, J. T.
Okereke, S. A.
Pearlman, M. R.
Porter, R. P.
Purdy, R. J.
Queenan, J. F.
Rabiner, R.
Reznek, S. R.
Ricketts, A. W., Jr.
Sarraquigne, M. R.
Schlessinger, L.
Sehn, G. J.
Shawaker, D. R.
Shroff, M. N.
Shulman, L. M.
Solin, S. A.
Spira, P. M.
Stavn, M. J.
Stokowski, S. R.
Tomes, C. F.
Tripp, A. P., Jr.
Valby, L. R.
Wachtel, J. M.
Wawzonek, J. J.
Weingrad, J. S.
Weintraub, A. C.
Wende, C. D.
Winsor, N. K.
Yansen, D. E.
Zilles, S. N.

Assistants and Technicians

Aquinde, P.
Arnold, Jane B.
Atlas, H.
Babcock, E. B.
Barrett, J. W.
Barrows, F. W.
Beaton, Catherine M.
Bella, C. J.
Berg, A. E.
Butler, R. E., Jr.
Carbone, Marie L.
Chance, Eleanor K.
Connolly, J. T.
Cook, J. F.
Coyle, J. E.
Cranmer, R. E.
Crist, F. X.
DiPietro, P. J.
Donahue, P. D.
Engler, R. R.
Fitzgerald, E. W., Jr.
French, Marjorie A.
Gay, H. D.
Grande, Esther D.
Greenwood, E. L.
Gregor, C. A.
Hallett, J. G.
Karas, P.
Kelly, M. A.
Kierstead, J. D.
Lewis, R. R.
Major, Diane
McKenzie, J. A.
McLean, J. J.
Misail, M. L.
Molin, A. H.
Neal, R. W.
Nolund, J. F.
North, D. K.
Overslvizen, T.
Papa, D. C.
Peck, J. S.
Porcaro, R. J.
Pyle, Cynthia M.
Samson, P. R.
Schwabe, W. J.
Sears, A. R.
Smith, Gabriella W.
Sprague, L. E.
Stroud, Marion B.
Thompson, D. S.
Tortolano, A. J.
Yaffee, M. A.
Yee, F. Q.

Document Room

Hewitt, J. H.
Hurvitz, Rose S.

Drafting Room

Navedonsky, C. P., Foreman
Donahue, J. B.
Morss, Dorothy H.
Porter, Jean M.
Rollins, I. E.

PERSONNEL

Machine Shop

Keefe, J. B., Foreman
Barnet, F. J.
Bletzer, P. W.
Brennan, J.
Bunick, F. J.
Cabral, M., Jr.
Carter, C. E.
Daniels, W. M.
Gibbons, W. D.
Liljeholm, F. H.
MacDonald, J. R.
Marshall, J. J.
Muse, W. J.
Reimann, W.
Ryan, J. F.
Sanroma, J. B.
Shmid, E.
Smart, D. A.
Tucker, C. L.

Secretaries

Amidon, Willa T.
Barron, Gladys G.
Bertozzi, Norma
Blais, Gisele G.
Boyajian, Judith A.
Brown, Eileen A.
Brunetto, Deborah A.
Canniff, Kathleen A.
Carbone, Angelina
Cavanaugh, Mary C.
Chapman, Carol A.
Daly, Marguerite A.
Dordoni, Joan M.
Epstein, Elinor F.
Geller, Elaine J.
Gordon, Linda S.
Greulach, Vicki E.
Hall, Nancy K.
Healy, Sylvia K.
Imbornone, Elaine C.
Johnson, Barbara A.
Kaloyanides, Venetia
Keim, Norma R.
Lannoy, Doris E.
Laurendeau, Carole A.
Litchman, Sandra H.
Loeb, Charlotte G.
May, Nancy A.
McEntee, Doris C.
Milan, Marilyn A.
Morneault, Beverly A.
Morneault, Diane M.
Omansky, Betsey G.
Petone, Rosina C.
Rose, Martha G.
Scalleri, Mary B.
Scanlon, Dorothea C.
Smith, Claire F.
Solomon, Cynthia
Staffiere, Rose Carol
Thomson, Susan M.
Toebes, Rita K.
Touchette, Thelma B.
Townley, Madeline S.
Vesey, Patricia A.

Technical Typists

Barnes, R. A.
Fleming, Patricia L.
Levin, L.
Levine, R. I.
Rabkin, W. I.

Stock Rooms

Doiron, E. J., Foreman
Audette, A. G.
Haggerty, R. H.
Joyce, T. F.
Legier, D. O.
Lucas, W. G.
McDermott, J. F.
Massey, L. N.
Riley, J. F.
Sharib, G.
Sincuk, J., Jr.

Technicians' Shop

Lorden, G. J., Foreman
D'Amico, C. R.
Fownes, Marilyn R.
Howell, W. B.
Lander, H. J.
MacDonald, K. B.

Tube Laboratory

Staff

Rosebury, F.
Ryan, L. W.

Glass Blowers

DiGiacomo, R. M.
Doucette, W. F.

Technicians

Aucella, Alice A.
Griffin, J. L.
Leach, G. H., Jr.
MacDonald, A. A.

PUBLICATIONS AND REPORTS

MEETING PAPERS PRESENTED

Rice University, Electrical Engineering Graduate Seminar, Houston, Texas
December 3, 1962

P. A. Willis, Computer Utilization in Biomedical Engineering Problems (invited)

Symposium on Neural Modelling (sponsored by the Librascope Division of General Precision, Inc., and the Air Force Office of Scientific Research), Ojai, California
December 4-6, 1962

B. G. Farley, Simulation of Neuron Networks by Digital Computers (invited)

L. Stark, Retinal Operations in Accommodative Tracking – Non-linear Describing Function Approach (invited)

Texas Medical Center, University of Texas, Houston, Texas
December 4-7, 1962

P. A. Willis, Communication Characteristics of Nerve (invited)
Neurological Control Systems (invited)

Stanford Research Institute, Stanford, California.
December 5, 1962

L. Stark, Automatic Adaptive Pattern Recognition in Electrocardiograms (invited)

Basic Sciences Group of Boston Section of AIEE, Boston, Massachusetts
December 18, 1962

W. T. Peake, Research on the Auditory Nervous System – The Role of Engineers (invited)

American Association for the Advancement of Science, Biometrics Division, Philadelphia, Pennsylvania
December 26, 1962

M. Eden, A Critical View of Computers in Medical Diagnosis (invited)

Modern Language Association Meeting, Washington, D. C.
December 27, 1962

E. S. Klima, Relatedness between Grammatical Systems (invited)

American Physical Society Winter Meeting in the West, Stanford University, Stanford, California

December 27-29, 1962

W. M. Whitney and C. E. Chase, Ultrasonic Dispersion in Liquid Helium below 1°K

Annual Meeting, Linguistic Society of America, New York

December 28, 1962

P. M. Postal, The Formal Characteristics of Grammatical Agreement

Woods Hole Oceanographic Institute, Woods Hole, Massachusetts

January 7, 1963

L. Stark, On-line Digital Computation in Neurological Experiments (invited)

American Chemical Society Meeting, Cincinnati, Ohio

January 16, 1963

G. G. Hammes, Studies of Fast Reactions in Enzymatic Systems

New York Meeting, American Physical Society, New York

January 23-26, 1963

J. F. Cochran and M. Yaqub, Magneto-resistance Effects in Small Gallium Single Crystals

D. H. Douglass, Jr., M. S. Dresselhaus, and R. L. Kyhl, Measurement of the Magnetic Field Dependence of the Surface Impedance of Superconducting Tin at 9 kmc

W. D. Jackson, C. R. Phipps, and R. A. Reitman, A-C Properties of Superconducting Wires and Ribbon

C. A. Shiffman, M. Garber, J. F. Cochran, E. Maxwell, and G. W. Pearsall, The Specific Heats of Tin-Lead Alloys

W. W. Smith, Hyperfine Structure and Magnetic Moments of Hg^{195} and Hg^{195*} in the 6^3P_1 State by Level Crossing

S. H. Wemple, Electric Field Effects on EPR Spectra in $KTaO_3:Fe^{3+}$

Association for Symbolic Logic Meeting, University of California, Berkeley

January 26, 1963

T. More, Jr., Calculi of Sequences of Propositional Formulas
On the Modification of Modus Ponens

Two Lectures, Université Catholique de Louvain, Belgium

February 4, 1963

Y. W. Lee, Statistical Theory of Nonlinear Systems (invited)

The Seventy-Fifth Anniversary Symposium on Engineering for Major Scientific Programs, Georgia Institute of Technology, Atlanta, Georgia

February 5-6, 1963

W. A. Rosenblith, Computers and Brains: Competition and/or Co-existence (invited)

Third Quantum Electronics Conference, Paris

February 11-15, 1963

G. W. Stroke, Theory, Production, and Use of Optical Gratings for High-Resolution Spectroscopy

Conference on Sensory Evoked Response in Man (sponsored by The New York Academy of Sciences, Division of Biological and Medical Sciences, and Albert Einstein College of Medicine), New York

February 14-15, 1963

J. S. Barlow, Evoked Response in Relation to Visual Perception and Reaction Times in Man

Biophysical Society Seventh Annual Meeting, New York

February 18-20, 1963

B. G. Farley, Similarities between the Behavior of a Neuron-like Network Model and "Slow-Wave" Phenomena on the Brain (invited)

J. L. Hall II, Binaural Interaction in Single Units of Accessory Superior Olivary Nucleus of Cats

L. Stark, Control of Pupillary and Eye Movements (invited)

A. Troelstra, A Non-linear Model for the Calculation of the DC Component in the Human Electro-retinogram (ERG)

T. F. Weiss, A Model for Firing Patterns of Auditory Nerve Fibers

1963 International Solid-State Circuits Conference, University of Pennsylvania, Philadelphia, Pennsylvania

February 20-22, 1963

J. Y. Lettvin, Considerations Underlying the Study of Sensory Elements (invited)

MEETING PAPERS PRESENTED (continued)

Seminar, Department of Electrical Engineering, Johns Hopkins University, Baltimore, Maryland

February 25, 1963

M. H. Goldstein, Jr., The Role of Hearing in Animal Communication (invited)

Symposium on Metallurgy of Electron Emitting Materials, Dallas, Texas

February 25, 1963

W. B. Nottingham, Electron Emitting Materials

JOURNAL ARTICLES ACCEPTED FOR PUBLICATION

(Reprints, if available, may be obtained from the Document Room, 26-327, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.)

M. A. Arbib, Monogenic Normal Systems Are Universal (J. Australian Math. Soc.)

D. E. Baldwin, Close Collisions in a Plasma (Phys. Fluids)

J. S. Barlow, R. L. Rovit, and P. Gloor, Correlation Analysis of EEG Changes Induced by Unilateral Intracarotid Injection of Amobarbital (EEG Clin. Neurophysiol.)

J. S. Bross, Problems of Equivalence in Some German and English Constructions (MT)

S. C. Brown, G. Bekefi, and R. E. Whitney, Far Infrared Interferometer for the Measurement of High Electron Densities (J. Opt. Soc. Am.)

D. H. Douglass, Jr. and M. W. P. Strandberg, Stages in the Education of a Physicist: An Attempted Solution of a Pedagogical Problem (Am. J. Phys.)

R. M. Fano, A Heuristic Discussion of Probabilistic Decoding (IRE Trans., PGIT)

M. C. Goodall, Realizability of Inductive Logic (IRE Trans., PGME)

J. L. Hirshfield and G. Bekefi, Decameter Radiation from Jupiter (Nature)

J. J. Katz and J. A. Fodor, The Structure of a Semantic Theory (Language)

S. J. Keyser, Review of "The Pronunciation of English in the Atlantic States" by Hans Kurath and Raven I. McDavid, Jr. (Language)

K. Kornacker, Some Properties of the Afferent Pathway in the Frog Corneal Reflex (Exptl. Neurol.)

C. L. Liu, A Property of Partially Specified Automata (Information and Control)

D. J. Sakrison, Iterative Design of Optimum Filters for Non Mean-Square Error Performance Criteria (IRE Trans., PGIT)

JOURNAL ARTICLES ACCEPTED FOR PUBLICATION (continued)

J. I. Steinfeld and G. G. Hammes, The Relaxation Spectrum of Nickel-Triglycine Complexes (J. Phys. Chem.)

K. N. Stevens, T. T. Sandel, and A. S. House, Perception of Two-Component Noise Bursts (J. Acoust. Soc. Am.)

M. W. P. Strandberg, Book Review of "Elementary Solid State Physics: A Short Course" by Charles Kittel (Proc. IRE)

M. W. P. Strandberg, IRE Millimeter and Submillimeter Conference, Orlando, Florida, January 7-10, 1963 (Appl. Optics)

G. W. Stroke, Interferometric Method of Velocity of Light Measurement (Appl. Optics)

L. Tisza and P. M. Quay, The Statistical Thermodynamics of Equilibrium (Ann. Phys.)

LETTERS TO THE EDITOR ACCEPTED FOR PUBLICATION

F. T. Barath, A. H. Barrett, J. Copeland, D. E. Jones, and A. E. Lilley, Preliminary Results of the Mariner II in Microwave Radiometer Experiment (Science)

G. Fiocco and E. Thompson, Thomson Scattering of Optical Radiation from an Electron Beam (Phys. Rev. Letters)

P. Penfield, Jr. and B. L. Diamond, Letter to the Editor (concerning varactor multipliers) (Microwave J.)

P. Penfield, Jr. and R. P. Rafuse, Frequency Multipliers and Harmonic Generators Using Varactor Diodes (Semiconductor Products and Solid State Technology)

P. Penfield, Jr. and R. P. Rafuse, List of Corrections for "Varactor Applications" (The M.I.T. Press, Cambridge, Mass., 1962) (Microwave J.)

W. J. Tomlinson III and H. H. Stroke, Hyperfine Structure, Isotope and Isomer Shifts in Hg^{193m}, Hg^{194}, Hg^{195m}, Hg^{195} – Comments on the Paper of Kleiman and Davis (J. Opt. Soc. Am.)

TECHNICAL REPORTS PUBLISHED

(These and previously published technical reports, if available, may be obtained from the Document Room, 26-327, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.)

408 William H. Heiser, Axial Field Effects in a Magnetically Driven Shock Tube

409 Laszlo Tisza, The Conceptual Structure of Physics

410 James L. Massey, Threshold Decoding

SPECIAL TECHNICAL REPORTS PUBLISHED

(These monographs are available from The M. I. T. Press, 238 Main Street, Cambridge 39, Massachusetts.)

Number 8 W. P. Allis, S. J. Buchsbaum, and A. Bers, Waves in Anisotropic Plasmas

SPECIAL PUBLICATIONS

W. P. Allis, Wave Propagation in Anisotropic Plasmas (Proc. Symposium on Electromagnetic Theory and Antennas, The Technical University of Denmark, Copenhagen, June 25-30, 1962)

J. S. Barlow, Evoked Response in Relation to Visual Perception and Reaction Times in Man (Ann. New York Acad. Sci.)

G. O. Barnett and W. D. Jackson, Measurement of Instantaneous Blood Flow (Progress in Cardiovascular Diseases)

A. H. Barrett, Microwave Spectral Lines as Probes of Planetary Atmospheres (Mémoires de la Société Royale des Sciences de Liège)

W. S. McCulloch, A Historical Introduction to the Postulational Foundations of Experimental Epistemology (a chapter in Epistemology in Anthropology, edited by Helen L. Livingston)

R. Melzack, Effects of Early Experience on Behavior: Experimental and Conceptual Considerations (Psychopathology of Perception, edited by P. Hoch and J. Zubin, American Psychopathological Association, New York)

Introduction

This report, the sixty-ninth in a series of quarterly reports issued by the Research Laboratory of Electronics, contains a review of the research activities of the Laboratory for the three-month period ending February 28, 1963. Since this is a report on work in progress, some of the results may not be final.

RADIO PHYSICS

I. MICROWAVE SPECTROSCOPY*

Prof. M. W. P. Strandberg	R. Huibonhoa	H. Pauwels
Prof. R. L. Kyhl	J. G. Ingersoll	Mahin Rahmani
Dr. B. D. N. Rao	P. F. Kellen	W. J. Schwabe
J. M. Andrews, Jr.	J. D. Kierstead	J. R. Shane
J. C. Burgiel	S. H. Lerman	N. Tepley
Y. H. Chu	J. W. Mayo	C. F. Tomes

A. ULTRASONIC ATTENUATION IN SUPERCONDUCTORS

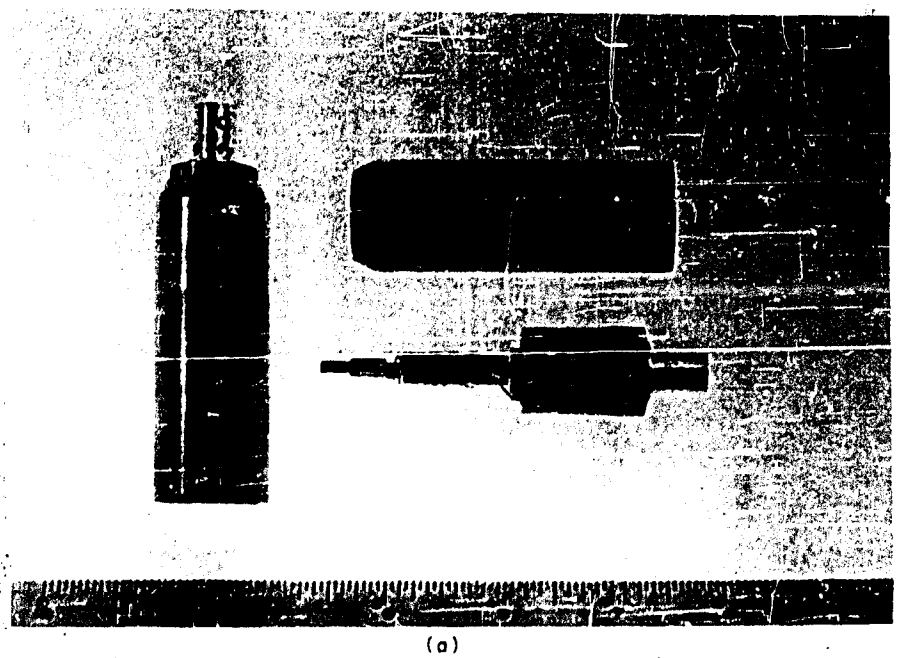
The work described here was initiated several months ago in the very high frequency range. Equipment was designed and constructed for the generation and reception of ultrasonic pulses at 165 mc. The transmitter consists of the high-frequency and blocking-oscillator portions of the circuit used by Chick, Anderson, and Truell.[1] The transmitting and receiving transducers are identical helical resonators[2] exciting x-cut quartz rods that are inserted in an rf electric field produced in a gap of nearly identical geometry to that of the re-entrant cavity used in the microwave phonon experiments.[3] These helical resonators are shown in Fig. I-1a. The two quartz rods (0.118 inch in diameter, 0.504 inch long) with a metal specimen (usually approximately 0.1 inch thick) sandwiched between their ends are inserted through a small hole in the end of each helical resonator into the electric-field gap. The resonators are then clamped in this coaxial configuration, mounted on stainless-steel coaxial lines, and immersed in a standard double-dewar helium cryostat. Signal power from the receiving transducer is mixed with local oscillator power in a strip-line hybrid coupler; the dual output is then passed through a balanced detector to a 30-mc, 120-db, if strip.

This apparatus has been used to study ultrasonic attenuation in specimens of indium and mercury, both in the superconducting and normal states. The metals were obtained from standard chemical reagent stock (In, 99.99 per cent pure; Hg, 99.999 per cent pure). No attempt was made to obtain single crystals with known orientation. In the case of indium, the quartz transducer rods were clamped in a vee-block and the indium was allowed to solidify in a Teflon mold.[4] The mercury specimen was contained in a small, cylindrical aluminum capsule; the quartz transducer rods had been fitted into aluminum bushings that were inserted into the ends of the capsule and cemented into place. (See Fig. I-1b.)

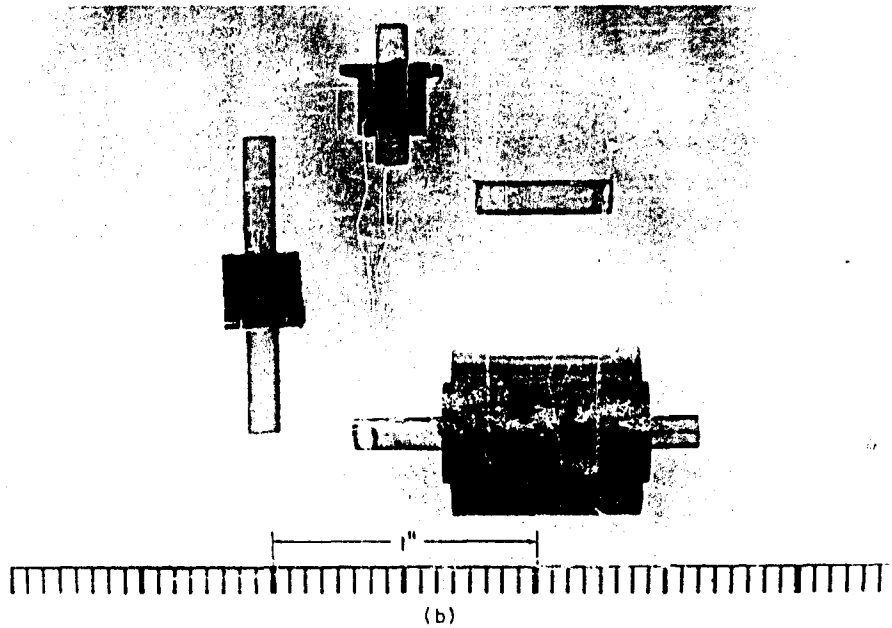
1. Experiment

In a typical experiment the entire transducer assembly containing the metal specimen is cooled to some fixed temperature well below the superconducting transition

*This work was supported in part by Purchase Order DDL B-00368 with Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology with the joint support of the U. S. Army, Navy, and Air Force under Air Force Contract AF19(604)-7400.

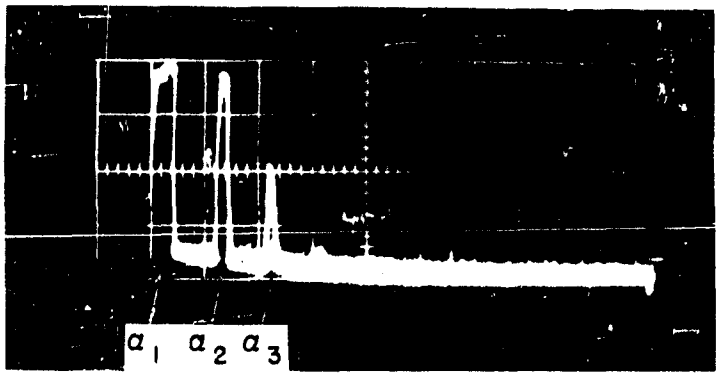


(a)

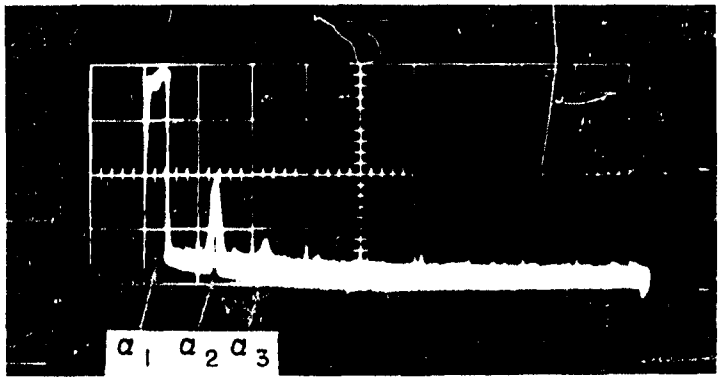


(b)

Fig. I-1. (a) Helical resonators used for the transmitting and receiving transducers.
(b) Mercury capsule and the indium mold. The quartz transducer rods are shown protruding from each.



(a)



(b)

Fig. I-2. (a) Acoustic pulses transmitted through superconducting mercury. a_1 is rf leakage; a_2, an acoustic pulse that has passed through both quartz rods and the mercury specimen; a_3, superposition of two acoustic pulses, one having made a double reflection in the first rod, and the other, a double reflection in the second rod.
(b) Acoustic pulses transmitted through normal mercury. This is exactly the same situation as in (a), except that a magnetic field $H \gg H_c(T)$ has been switched on.

temperature. The transmitter initiates, in the transmitting transducer, a pulse of rf power of approximately 25 watts, peak, and 1 μsec in duration. A small fraction of this power ($\approx 10^{-3}$) is converted into a pulse of acoustic energy of longitudinal polarization by the piezoelectric effect. This pulse passes along the first quartz rod, through the specimen, and into the second quartz rod. At the receiving transducer the same fraction of the acoustic power is reconverted into an rf signal, and the rest is reflected. Figure I-2 shows the resulting oscilloscope traces. The first pulse is rf leakage. The second is an acoustic pulse that has passed through both quartz rods and the metal specimen. The third pulse is actually a superposition of two acoustic pulses: one has made a double reflection in the first rod; the other, a double reflection in the second rod. The superposition is caused by the fact that the two quartz transducer rods are equal in length (within ±.001 in.).

An electromagnet surrounds the portion of the cryostat containing the metal specimen, which is capable of providing a variable magnetic field normal to the ultrasonic wave vector and the cylindrical axis of the metal specimen. If a magnetic field greater than the critical field of the superconductor is switched on, the metal returns to the normal state and the attenuation increases abruptly. Figure I-2b shows the same pulse-echo pattern as that shown in Fig. I-2a, except that the magnetic field $H \gg H_c(T)$ has been turned on. This change in the ultrasonic attenuation is due entirely to the effect of the conduction electrons in the metal.

The relative attenuation was measured as a function of applied magnetic field for both metals, and is plotted in Fig. I-3. Observe that it is a continuous function, a result that is implicit in the aluminum data of David, Van der Laan, and Poulis.[5] Data on the total change in attenuation between the normal and the superconducting states should probably be taken from the first pulse only. The reason for this is that the acoustic energy that arrives at the receiving transducer at the time indicated by the second pulse, since it is a superposition of two acoustic pulses that may well have travelled over slightly different paths, has an unknown phase factor determined by the details of the paths.

A specimen of mercury, approximately 0.25-in. thick, was prepared in order to obtain reasonably reliable velocity data. The velocity was measured at room temperature at 165 mc and found to be 1.47 ± .03 km/sec, which is in agreement with the measurements of Ringo, Fitzgerald, and Hurdle.[6] A specimen of this length, however, is more vulnerable to thermal contraction effects. Thus far, it has not been possible to maintain satisfactory acoustic contact between such a long specimen and the quartz transducer rods as the temperature is lowered.

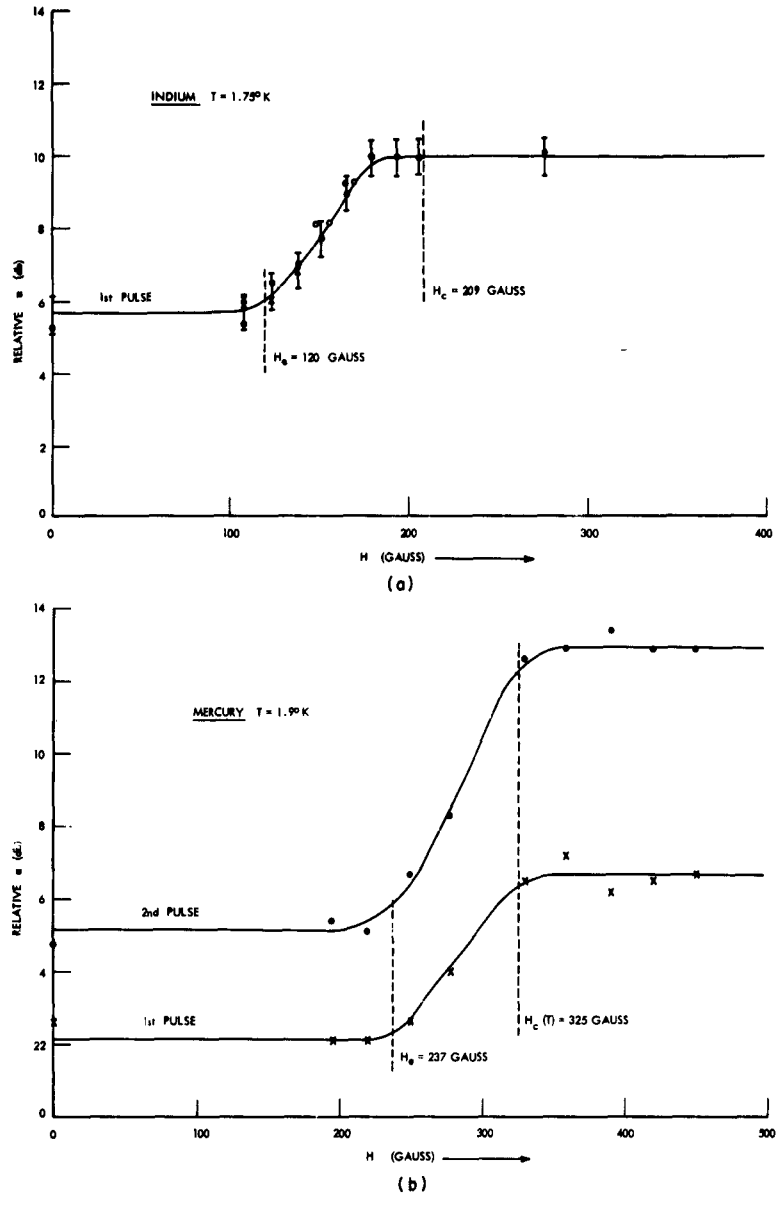


Fig. I-3. (a) Relative attenuation of a 165-mc ultrasonic pulse in indium as a function of magnetic field.
(b) Relative attenuation of 165-mc ultrasonic pulses in mercury as a function of magnetic field.

2. Discussion

An important mechanism for ultrasonic attenuation in metals at sufficiently low temperatures that the electronic mean-free path becomes of the same order of magnitude as the acoustic wavelength (usually below 10°K) is the electron-phonon scattering. Thus, if we have P_o acoustic power incident on a metal specimen of length L, the power P leaving the specimen will be

$$P = P_o \exp[-(\alpha_{elec}+\alpha_o)L]. \tag{1}$$

Here, α_{elec} is that portion of the attenuation coefficient caused by scattering from the conduction electrons; α_o is that portion of the attenuation coefficient caused by scattering from thermal phonons, lattice defects, grain boundaries, and so forth. The coefficient α_o is temperature-independent in the range of temperatures which we are considering, and therefore will not be discussed further. Using the free-electron model, Pippard has calculated expressions for the electronic contribution to the attenuation coefficient for arbitrary mean-free path.[7] This approach has been well justified by the experimental results of Morse at frequencies between 9 mc and 56 mc.[8] Pippard's result for longitudinal polarization is

$$\alpha_n = \alpha' \cdot \frac{6}{\pi}\left(\frac{1}{ql}\right)\left[\frac{1}{3}\frac{(ql)^2 \tan^{-1}(ql)}{ql - \tan^{-1}(ql)} - 1\right], \tag{2}$$

where

α_n ≡ the electronic contribution to the attenuation coefficient for the metal in the normal state

q ≡ the acoustic wave vector

l ≡ the electronic mean-free path.

The coefficient α', which is the limiting form of α_n for long electronic mean-free path, $ql \gg 1$, is given by

$$\alpha' = \frac{\pi N m v_o \omega}{6\rho_o u_\ell^2}, \tag{3}$$

where

N ≡ the number of free electrons per unit volume

m ≡ the electronic mass

v_o ≡ velocity of the free electrons at the Fermi surface

ω ≡ the ultrasonic frequency

ρ_o ≡ the density of the metal specimen

u_ℓ ≡ the longitudinal acoustic velocity in the metal specimen.

Reasonable estimates based upon the free-electron model, under the assumption of

one free electron per atom, yield the following values for α' at 165 mc:

Indium $\alpha' = 4.2\ \text{cm}^{-1}$

Mercury $\alpha' = 8.6\ \text{cm}^{-1}$.

The Bardeen-Cooper-Schrieffer (BCS) theory of superconductivity has been applied to the calculation of the ultrasonic attenuation in metals in the superconducting state.[9] Their result is

$$\frac{\alpha_s}{\alpha_n} = 2f[\epsilon(T)], \tag{4}$$

where

α_s = the electronic contribution to the attenuation coefficient for the metal in the superconducting state

$f[\epsilon]$ = the Fermi function of the superconducting energy gap

$\epsilon(T)$ = the temperature-dependent superconducting energy gap.

Let us represent the change, in decibels, of the ultrasonic attenuation coefficient between the normal and the superconducting states at any given temperature by D(T). Then

$$\alpha_n = \frac{D(T)}{10 \log_{10}(e)\, L \tanh\left[\frac{\epsilon(T)}{2kT}\right]} \tag{5}$$

where

L = the length of the specimen in cm

k = Boltzmann's constant

T = the temperature in °K.

By using the theoretical curve for $\epsilon(T)$ given in BCS,[9] it is possible to determine α_n from our experiments on indium and mercury.

Specimen	Purity (per cent)	D(T) (db)	T (°K)	L (cm)	$(\alpha_n)_{exp}$ (cm^{-1})
In	99.99	3.7 (±.5)	1.75	.172	5.0
Hg	99.999	5 (±1.0)	1.9	.252	4.6

These α_n, determined experimentally, are of the same order of magnitude as the limiting values α' calculated, by using Eq. 3, for the metals on the basis of the free-electron model. We conclude that for 165 mc the purities of the metals are sufficient to provide electronic mean-free paths at least of the same order of magnitude as the acoustic wavelength. That is, $ql \geq 1$.

As we have already pointed out, the relative attenuation, plotted as a function of magnetic field, does not yield a discontinuity at H_c, but rather begins a slow rise at H_e, well below H_c. This might be due to the presence of metallic impurities. Superconducting alloys are sometimes characterized by such a slow transition as a function of magnetic field, but this generally occurs at fields greater than H_c for the corresponding pure elements. A more cogent explanation is that the metal enters the intermediate state at H_e, and then proceeds to the normal state in a continuous manner, the transition reaching completion at $H_c(T)$. This intermediate state is a property of macroscopic superconductors, and depends upon the geometry of the specimen. The same behavior has been observed in the resistivity of macroscopic superconductors as a function of magnetic field.[10] The temperature dependence of H_c is given by

$$H_c(T) = H_c(0)\left[1 - \left(\frac{T}{T_c}\right)^2\right]. \tag{6}$$

This formula is an approximation, but it is sufficiently accurate for our use. Here, $H_c(0)$ is the critical magnetic field at absolute zero, and T_c is the critical temperature at zero magnetic field. Both are properties of the particular superconductor under investigation. We now evaluate $H_c(T)$ for each specimen using their critical constants and the temperature, which was determined from the vapor pressure of the liquid helium.

Specimen	$H_c(0)$ (gauss)	T_c (°K)	T (°K)	$H_c(T)$ (gauss)	H_e (gauss)	n_{exp}
In	283	3.41	1.75	209	120	.43
Hg	411	4.15	1.9	325	237	.27

H_e was determined from the graphs of the experimental data (Fig. I-3).

The macroscopic theory of superconductivity predicts that the metal will enter the intermediate state at a field H_e that is such that

$$\frac{H_e}{H_c} = 1 - n. \tag{7}$$

This formula holds for macroscopic bodies whose dimensions are large compared with a characteristic length $\Delta \approx 10^{-5}$ cm. The product $4\pi n$ is the demagnetizing coefficient of the specimen which takes into account the effect of its magnetization on the external magnetic field while in the superconducting state. Obviously, n depends upon the shape of the specimen. We list a few theoretical n values for simple cases:

	n
Sphere	1/3
Long cylinder in transverse field	1/2
Long cylinder in parallel field	0
Large disc with its plane normal to field	1

The shapes of our specimens were not controlled accurately, but can be described approximately. The indium was formed into a disc, 0.30 cm in diameter and 0.172 cm thick. At one edge of the disc, however, there was a sprue almost as large as the disc. The applied magnetic field was oriented parallel to the plane of the disc. The mercury was contained in a cylinder, 0.635 cm in diameter and 0.635 cm long. The quartz rods penetrated 0.192 cm into the specimen at each end. The applied magnetic field was perpendicular to the axis of the cylinder. The experimental values for n show, indeed, that the shape of the specimen is important in determining the onset of the intermediate state.

The author wishes to express his appreciation to E. C. Ingraham for the construction of the helical resonators, the mercury capsule, and portions of the helium cryostat.

J. M. Andrews, Jr.

References

1. B. Chick, G. Anderson, and R. Truell, J. Acout. Soc. Am. 32, 186 (1960).

2. W. W. Macalpine and R. O. Schildknecht, Proc. IRE 47, 2099 (1959).

3. P. H. Carr and M. W. P. Strandberg, J. Phys. Chem. Solids 23, 923 (1962).

4. The author is indebted to Norman Tepley for the fabrication of the indium specimen.

5. R. David, H. R. Van der Laan, and N. J. Poulis, Physica 28, 330 (1962).

6. G. R. Ringo, J. W. Fitzgerald, and B. G. Hurdle, Phys. Rev. 72, 87 (1947).

7. A. B. Pippard, Phil. Mag. 46, 1104 (1955).

8. R. W. Morse, Progress in Cryogenics, Vol. 1, edited by K. Mendelssohn (Heywood and Company, Ltd., London, 1959), p. 221.

9. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. 108, 1175 (1957).

10. W. J. de Haas, J. Voogd, and J. M. Jonker, Physica 1, 281 (1934).

II. MICROWAVE ELECTRONICS*

Prof. L. D. Smullin	Prof. L. J. Chu	R. P. Porter
Prof. H. A. Haus	R. J. Briggs	H. M. Schneider
Prof. A. Bers	P. A. Mandics	R. S. Smith

A. WAVES IN A BRILLOUIN-FOCUSED ELECTRON-BEAM WAVEGUIDE

The only known wave solution for the Brillouin-focused[1] electron beam are the ϕ-independent slow waves that were found from an approximate (quasi-static and weak space-charge) analysis.[2] We shall give here the exact solution for all waves.

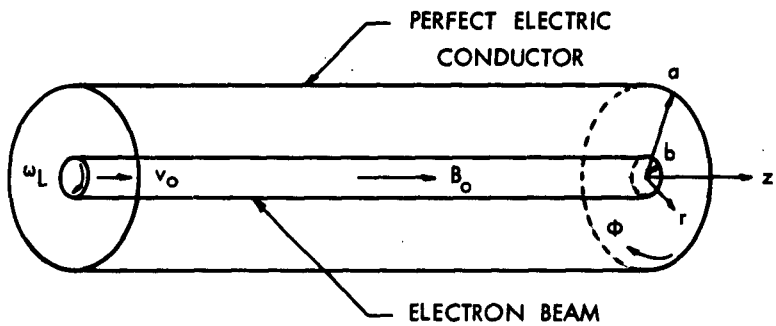


Fig. II-1. Brillouin-focused, circular electron-beam waveguide.

Consider a nonrelativistic Brillouin-focused electron-beam waveguide (Fig. II-1). In the unperturbed state this beam is characterized by a uniform velocity v_o along the applied magnetic field B_o, a constant angular velocity equal to the Larmor frequency $\omega_L = -eB_o/2m$, and the Brillouin-focusing condition

$$\omega_L^2 = \frac{1}{2}\omega_p^2, \tag{1}$$

where $\omega_p^2 = e\,\rho_o/m\epsilon_o$ is the plasma frequency.

The small-signal equations of this system are most conveniently written in a coordinate system that is translating with velocity v_o and rotating with angular velocity ω_L. Quantities relating to this coordinate system will be designated by double primes. In this frame Maxwell's equations take the simple form

$$\nabla \times \bar{E}'' = -j\omega\mu_o \bar{H}'' \tag{2}$$

*This work was supported in part by Purchase Order DDL B-00368 with Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology with the joint support of the U. S. Army, Navy, and Air Force under Air Force Contract AF19(604)-7400.

$$\nabla \times \bar{H}'' = j\omega\epsilon_o \left(1 - \frac{\omega_p^2}{\omega''^2}\right) \bar{E}'' \tag{3}$$

which corresponds exactly to the equations that govern an isotropic plasma waveguide. The complete solutions for waves in an isotropic plasma waveguide are known.[3] Hence the complete solutions for the waves in a Brillouin-focused electron-beam waveguide can readily be obtained by a proper transformation of the isotropic plasma waveguide solutions.

1. Dispersion Characteristics

The dispersion characteristic for the isotropic plasma waveguide is shown[3] in Fig. II-2. Here we shall consider only those branches that contain slow waves ($\beta > \omega/c$). For $\omega_p < \omega_{co}$ (the cutoff frequency of the empty waveguide) the branches that contain only fast waves occur for $\omega > \omega_p$, and are not shown in Fig. II-2. The dispersion characteristics shown are for the field solutions with dependence $\exp j(\omega''t'' - m''\phi'' - \beta''t'')$.

The dispersion characteristics in a coordinate system (denoted by single primes) that is rotating at a constant angular velocity ω_L are readily obtained through the nonrelativistic transformations

$$\omega'' = \omega' - m'\omega_L \tag{4}$$

$$m'' = m' \tag{5}$$

$$\beta'' = \beta'. \tag{6}$$

The results are shown in Fig. II-3. For most electron-beam systems of interest, $r\omega_L \ll c$ and the nonrelativistic transformations of Eqs. 4 and 5 apply. The first-order corrections to these transformations can be obtained from general relativity,[4] and are

$$\omega'' \cong (\omega' - m'\omega_L)\gamma_\phi \tag{7}$$

$$m'' \cong \frac{m'}{\gamma_\phi}, \tag{8}$$

where

$$\gamma_\phi = \left(1 - \frac{r^2\omega_L^2}{c^2}\right)^{-1/2} \cong 1 + \frac{r^2\omega_L^2}{2c^2}. \tag{9}$$

The dispersion characteristics in the laboratory frame, which is translating at the constant velocity v_o with respect to the primed frame, are obtained through the nonrelativistic transformation

$$\omega' = \omega - \beta v_o \tag{10}$$

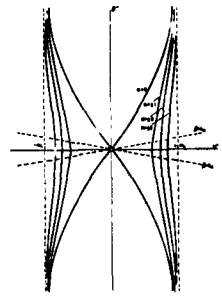


Fig. II-2. Dispersion characteristics in the rotating and translating frame-plasma frame.

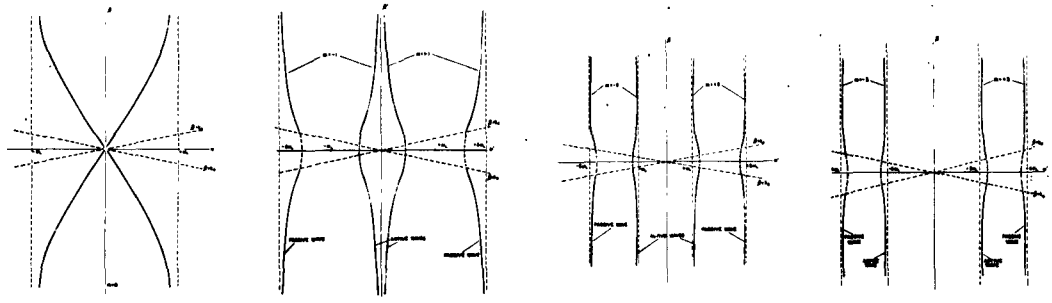


Fig. II-3. Dispersion characteristics in the frame rotating with respect to the plasma frame (translating with respect to the laboratory frame). (a) m = 0, (b) m = ±1, (c) m = ±2, (d) m = ±3.

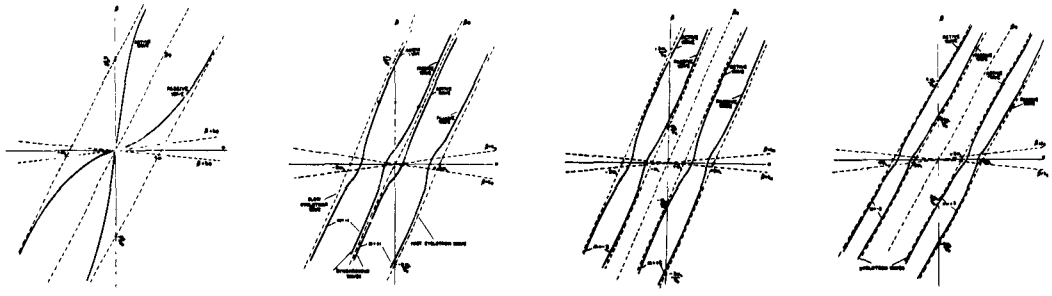


Fig. II-4. Dispersion characteristics in the laboratory frame. (a) m = 0 (no backward waves), (b) m = ±1, (c) m = ±2, (d) m = ±3.

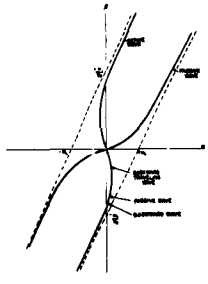


Fig. II-5. Dispersion characteristics in the laboratory frame for m = 0 waves containing a backward-wave region (Eq. 15).

$$\beta' = \beta. \tag{11}$$

The results are shown in Fig. II-4. The nonrelativistic transformation of Eqs. 10 and 11 is valid for $v_o \ll c$ and for slow waves, $\beta > \omega/c$. The relativistically correct transformations are readily obtained from special relativity,[4] and are

$$\omega' = (\omega - \beta v_o)\gamma_o \tag{12}$$

$$\beta' = \left(\beta - \frac{\omega v_o}{c^2}\right)\gamma_o, \tag{13}$$

where

$$\gamma_o = \left(1 - \frac{v_o^2}{c^2}\right)^{-1/2} \tag{14}$$

In Figs. II-4b-d we note that backward-wave characteristics (positive group velocity and negative phase velocity) are obtained for all waves with $m \neq 0$. For the waves with circular symmetry, $m = 0$, Fig. II-3a and Eq. 10 or 12 show that for certain beam velocities, v_o, backward-wave characteristics must also be possible. The condition for this possibility is

$$v_o < c\left[1 + \frac{I_o(k_p b)}{k_p b\, I_1(k_p b) \ln\frac{a}{b}}\right]^{-1/2}, \tag{15}$$

where I_o and I_1 are the modified Bessel functions, $k_p = \omega_p/c$, and c is the velocity of light in free space. Unlike the case for a confined flow beam, $B_o = \infty$, Eq. 15 is easily satisfied for most beams that can be produced in practice. Under the conditions of Eq. 15, the dispersion characteristic of Fig. II-4a changes to that shown in Fig. II-5.

2. Energy-Power Characteristics

The time-averaged, small-signal energy per unit length of waveguide, W_a'', in the isotropic plasma waveguide described by Eqs. 2 and 3, is positive definite.[3] Hence all of the waves in the double-prime frame are passive waves. The power flow in this frame, P_z'', is entirely electromagnetic.

In the coordinate system rotating at the angular frequency ω_L with respect to the double-prime frame, the nonrelativistic transformations of energy and power give[5]

$$\frac{W_a'}{\omega'} = \frac{W_a''}{\omega''} \tag{16}$$

$$\frac{P_z'}{\omega' v_g'} = \frac{P_z''}{\omega'' v_g''}, \tag{17}$$

where v_g is the group velocity (energy velocity) for the wave. The transformations in Eqs. 4 and 16 show how the active (negative energy) waves appear in the prime frame, Fig. II-3. The group velocity of the waves remains unchanged.

The transformations of energy and power to the laboratory frame[6] are

$$\frac{W_a}{\omega} = \frac{W'_a}{\omega'} \tag{18}$$

$$\frac{P_z}{\omega v_g} = \frac{P'_z}{\omega' v'_g} \tag{19}$$

The appearance of active waves can now be traced to the transformations in Eqs. 10 and 18. We note the existence of active waves with phase velocities that may be either slower or faster than the velocity of light and v_o.

A. Bers, R. S. Smith

References

1. L. Brillouin, A theorem of Larmor and its importance for electrons in magnetic fields, Phys. Rev. 67, 260-266 (1945).

2. W. W. Rigrod and J. A. Lewis, Wave propagation along a magnetically focused cylindrical electron beam, Bell System Tech. J. 33, 399-416 (1954); G. R. White, Space charge waves in relativistic Brillouin beams, Mikrowellenrohren, Vortrage der Internationalen Tagung in München, 7-11 June 1960; NTF, Vol. 22, pp. 271-278, 1961.

3. W. P. Allis, S. J. Buchsbaum, and A. Bers, Waves in Anisotropic Plasmas (The M. I. T. Press, Cambridge, Mass., 1963).

4. C. Møller, The Theory of Relativity (Oxford University Press, London, 1955).

5. R. J. Briggs, H. A. Haus, and A. Bers, Small-Signal Energy, Power and Momentum Relations, Internal Memorandum, Research Laboratory of Electronics, M. I. T., 1963.

6. P. A. Sturrock, In what sense do slow waves carry negative energy?, J. Appl. Phys. 31, 2052-2056 (1960).

III. MOLECULAR BEAMS

Prof. J. R. Zacharias
Prof. J. G. King
Prof. C. L. Searle
Dr. G. W. Stroke

R. S. Badessa
V. J. Bates
B. Brabson
R. Golub
G. L. Guttrich

W. D. Johnston, Jr.
S. G. Kukolich
F. J. O'Brien
C. O. Thornburg, Jr.

A. CESIUM BEAM TUBE INVESTIGATION

During the past quarter a new control system was investigated which has the desirable property of minimizing the effects of the transient behavior of the cesium beam tube. As we mentioned in Quarterly Progress Report No. 67 (page 45), the transient occurring in the beam tube when the X-band excitation frequency is modulated could make circuit phase tolerances so severe as to preclude obtaining a stability of one part in 10^{12}. The technique used in the new system consists essentially of utilizing square-wave frequency modulation of the excitation at a sufficiently low repetition rate (11 cps) to provide intervals that are comparatively free of transients, and then gating these transient-free intervals for use in the control loop.

The various waveforms generated in the system are shown in Fig. III-1. Figure III-1a is the 11-cps square wave from which all subsequent waveforms are derived. Its duty

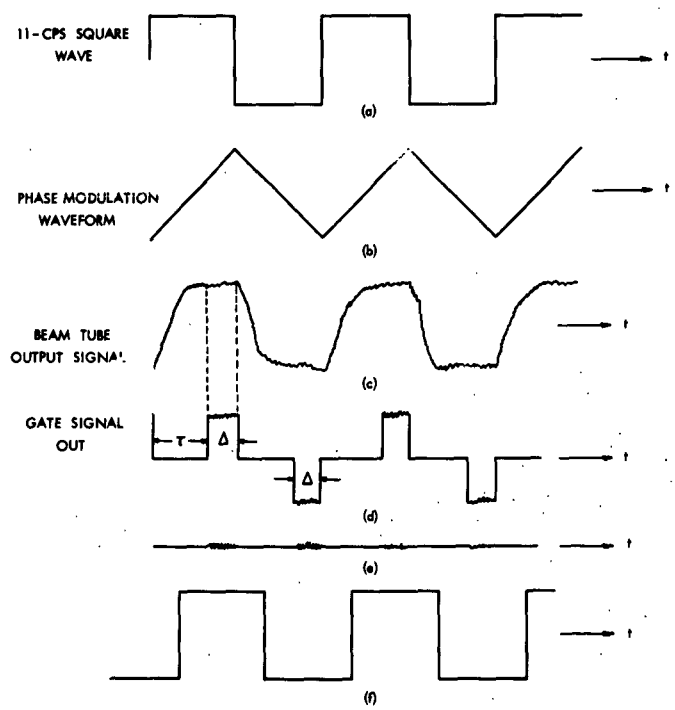


Fig. III-1. Waveforms present in the cesium clock system.

cycle of 50 per cent is guaranteed by driving a 2:1 flip-flop from a 22-cps pulse. The square wave, appropriately delayed, also serves as the reference for the synchronous detector of the loop. Figure III-1b is a triangular waveform obtained by integration of the square wave of Fig. III-1a. It is used for phase modulation of the 16.4-mc crystal oscillator that, by frequency multiplication to X-band, serves ultimately as the excitation signal. Triangular phase modulation conveniently yields the desired square-wave frequency-modulation waveform. Figure III-1c is the waveform observed at the output of the beam tube electron multiplier, when the average excitation frequency is located on one side of the cesium resonance peak. The exponential rise and decay is the transient, and represents a time constant approximately equal to the time of travel of the cesium atoms down the beam tube. It will be noted that the low repetition rate of 11 cps permits a relatively flat region to exist just before the next frequency excursion. It is this flat region that is gated for use in the loop, as is shown in Fig. III-1d. Figure III-1e shows how the fundamental component of the signal of Fig. III-1d disappears to form a 22-cps series of random noise bursts when the excitation frequency coincides exactly with that of cesium resonance. Figure III-1f is generated by delaying the signal of Fig. III-1a, and represents the reference signal for the synchronous detector. Figure III-2 is a block diagram of the complete system.

Figure III-3 shows the method used for checking the quadrature level. The synchronous detectors are mechanical choppers to prevent spurious direct current from being

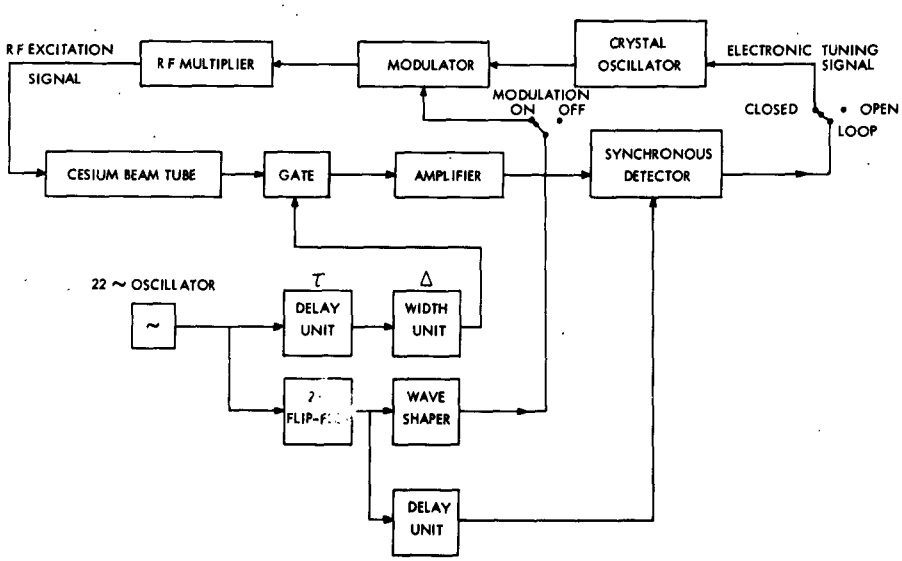


Fig. III-2. Cesium clock system utilizing time-domain filtering for minimizing transient effects.

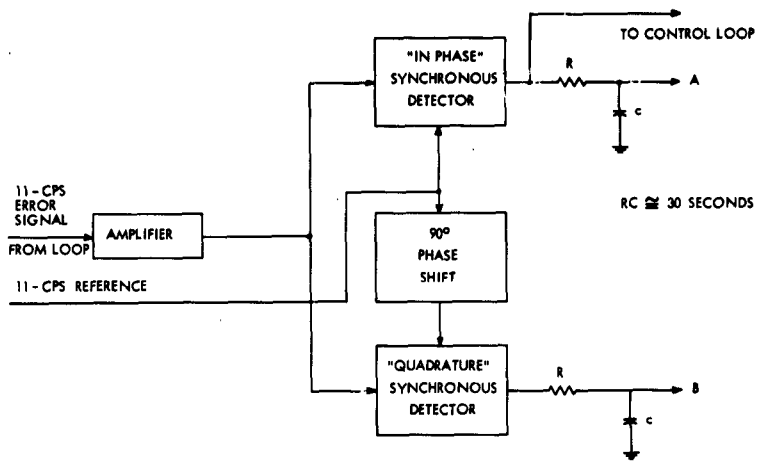


Fig. III-3. Double synchronous detectors for measurement of quadrature.

produced by reference unbalance. Because the references are 90° out of phase, the occurrence of zero direct current simultaneously at A and B can be achieved only by the complete absence of an 11-cps component in the incoming error signal. To obtain a calibration for the measurement, the loop is opened and the excitation frequency is detuned to one side of the resonance peak, to give maximum dc voltage at A. After measuring this voltage, the excitation frequency is recentered to the peak of the resonance, and the loop closed. Under the closed-loop condition the voltage at A will be very close to zero. The voltage at B will be a measure of quadrature and can be expressed as a fraction of the voltage that was observed at A under the open-loop condition. When measurements were made on the system of Fig. III-2, with a gate width Δ equal to one-third of a modulation half-period (see Fig. III-1d) and the gate was positioned just ahead of the next frequency excursion of the modulation, the quadrature level was well below that of the noise passing through the 30-sec time-constant filter. As the position of the gate was advanced toward the transient region, the quadrature appeared, as would be expected. The worst quadrature occurred when the gating was completely removed so that the signal to the synchronous detectors had the form shown in Fig. III-1c. Under this condition the quadrature was approximately 1.6 per cent of the in-phase signal level, or approximately 78 times the error voltage that would appear at the in-phase synchronous detector if the crystal oscillator were to detune one part in 10^{12}.

The stability of an atomic clock is determined partly by the stability of the resonance characteristic of its beam tube and partly by the stability of the electronics of the servo loop used to lock to it. These two effects are difficult to separate when seeking stabilities

of one part in 10^{12}. Certain performance requirements must be met by the loop, however, if such stability is to be achieved. For example, suppose that we were to remove the modulation from the excitation signal of the beam tube and open the loop (see Fig. III-2). Since the signal entering the synchronous detector should now be completely uncorrelated with the synchronous detector reference, no direct current should appear at its output. The random variations that do appear should not exceed the dc level that would occur when, in normal operation of the loop, the excitation frequency is detuned by one part in 10^{12}. The measurement is made with a filter having a time constant equal to the measurement time required for one part in 10^{12} on the basis of beam tube signal-to-noise ratio. In Fig. III-4 is shown a block diagram of the noise susceptibility measurement system. The results are shown in Fig. III-5.

In Fig. III-6 is shown a block diagram of a test performed to determine the interchangeability of two independently adjusted electronic systems. Atomic clock No. 1 was operated in the normal manner, but the 11-cps error signal appearing at the output of its electron multiplier was fed to the electronic apparatus of the other clock. Because of the servo action of the first clock's loop, very little 11-cps signal would be expected at the output of the electron multiplier. However, if the loop is not operating properly, for example, because of quadrature effects or overloading on noise peaks, an excessive amount of 11-cps signal could exist. The presence of this component would cause a dc output to appear at the synchronous detector of the second clock in

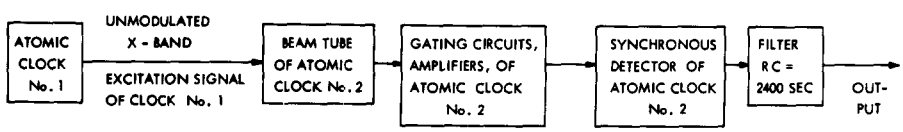


Fig. III-4. Measurement of susceptibility to noise.

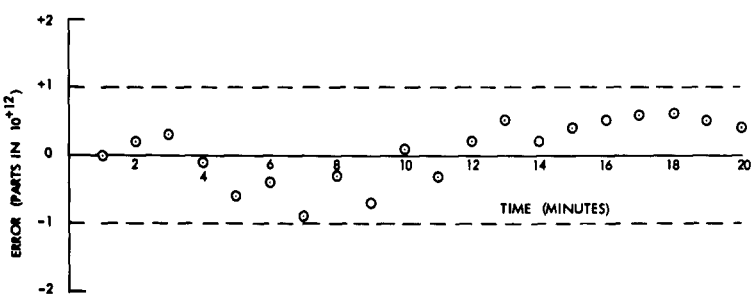


Fig. III-5. Noise susceptibility measurements.

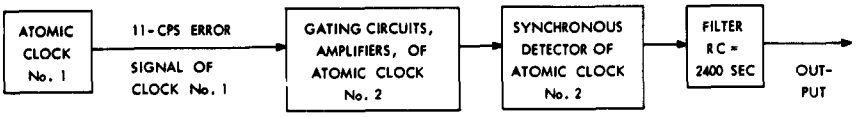


Fig. III-6. Measurement of interchangeability of two independently adjusted systems.

Fig. III-6. On the other hand, if the two electronic systems are in agreement and both are free of spurious effects, an input signal that gives a zero average out of one should give a zero average out of the other. Figure III-7 shows the results of such a test.

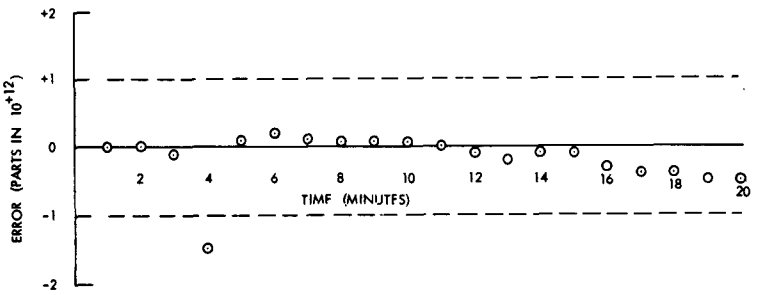


Fig. III-7. Electronics interchangeability measurements.

The two systems were adjusted independently and had in common only the 22-cps oscillator frequencies of Fig. III-2, necessitated by the fact that the synchronous detector of clock No. 2 was required to demodulate the modulation introduced by clock No. 1.

Measurements of frequency stability are now in progress. We hope to present the results of the tests in our next report.

R. S. Badessa, V. J. Bates, C. L. Searle

IV. RADIO ASTRONOMY*

Prof. A. H. Barrett
Prof. J. W. Graham
Prof. R. P. Rafuse
Dr. G. Fiocco
R. J. Allen
A. T. Anderson III
J. C. Blinn III
R. K. Breon
P. J. Brendel
G. A. Garosi
J. W. Kuiper
B. R. Kusse
D. H. Staelin
D. H. Steinbrecher

A. K-BAND RADIOMETRY AND OBSERVATIONS

Observations at 25.5 kmc were made of Venus, Taurus A, the Sun, and the Moon during the first two weeks of December 1962. These were made with the Lincoln

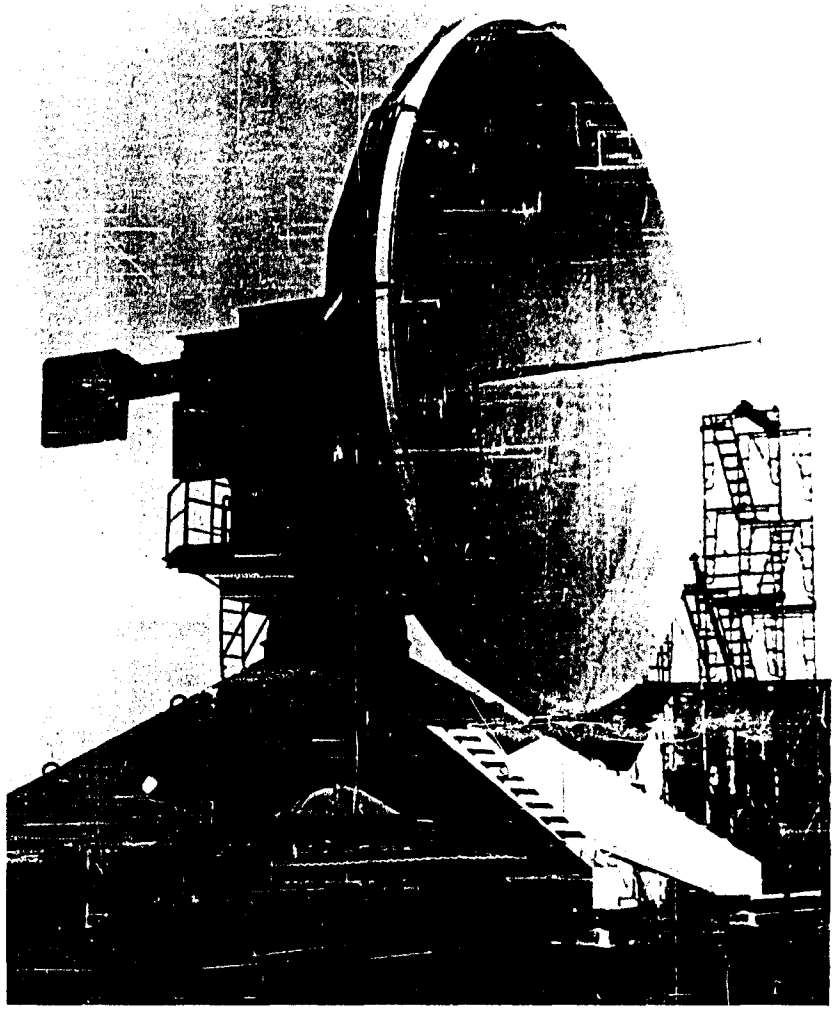


Fig. IV-1.
28-ft diameter parabolic antenna at Lincoln Laboratory, M.I.T.

*This work is supported in part by the National Aeronautics and Space Administration (Contract NaSr-101, Grant NsG-250-62, Grant NsG-264-62); and in part by the U.S. Navy (Office of Naval Research) under Contract Nonr-3963(02)-Task 2.

Laboratory 28-ft antenna shown in Fig. IV-1, and the superheterodyne radiometer described in Quarterly Progress Report No. 68 (pages 35-37). Observations were made simultaneously at 35.2 kmc at Lincoln Laboratory, M.I.T.

The data have not been completely processed, so only preliminary results are available.

The observations of Taurus, the Sun, and the Moon were all drift scans in which the antenna was pointed ahead of the source and held stationary while the source moved past. Venus was observed by using drift scans and on-off measurements. An on-off measurement consisted of tracking with the antenna pointed first at one side of the source, on the source, and then off again. Each of these three steps lasted one or two minutes.

Venus was observed with approximately 130 drift scans, Taurus A with 20, the Sun with 8, and the Moon with 9. Venus was also observed with approximately 60 on-off measurements. Brief attempts were made to observe Cassiopea A and Jupiter, but without success.

Figure IV-2 shows a typical drift scan of the sun. The variation in the center is believed to be due to small gain fluctuations, and the wings of the curve are due to the

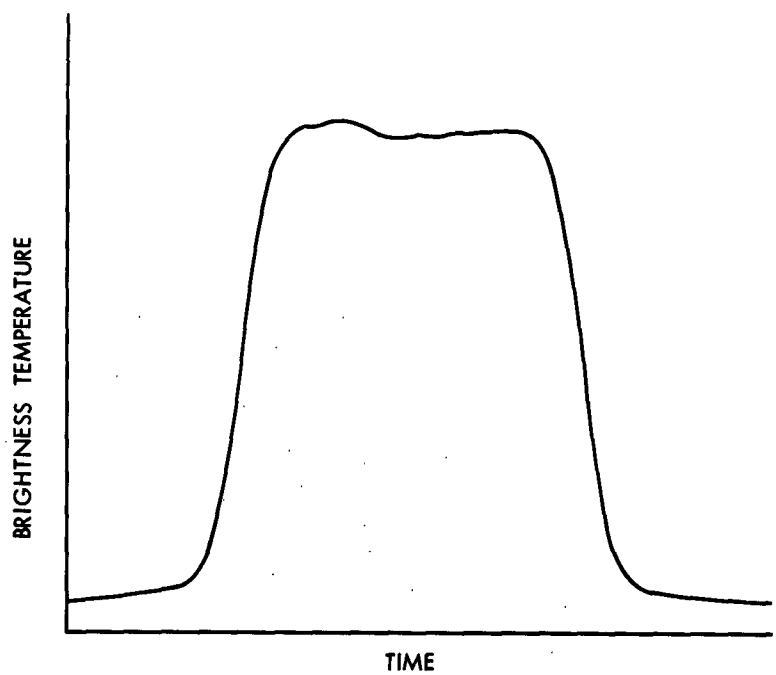


Fig. IV-2. Solar drift curve.

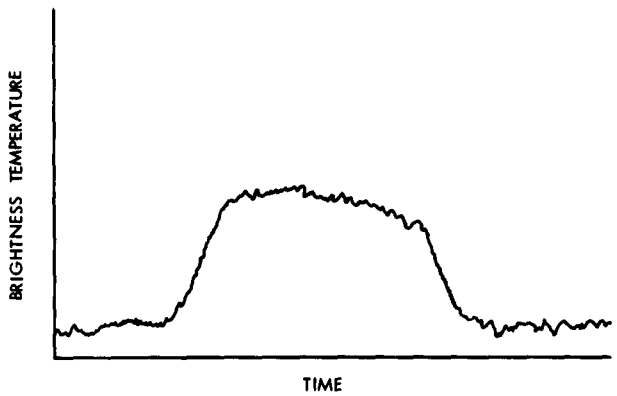


Fig. IV-3. Lunar drift curve.

antenna side lobes. Figure IV-3 shows a typical drift scan of the Moon. The asymmetry in the curve is due to the nonuniform temperature of the Moon's surface which exists 3.5 days before full Moon.

The preliminary results are:

1. Venus (on-off data), $\overline{T}_B = 400 \pm 100°K$
2. Taurus A (assuming disk of 3.6' diameter of uniform intensity),
   $S_\nu = 230(\pm 100) \times 10^{-26}$ w/m^{2}cps
3. Sun (center), $T_B = 8900 \pm 980°K$
4. Moon (center)
   3.5 days before full Moon, $T_B = 240 \pm 40°K$
   0.3 day before full Moon, $T_B = 254 \pm 30°K$
   1.8 days after full Moon, $T_B = 254 \pm 30°K$

In this list the error brackets are estimates.

These preliminary results are close to the expected values. The final results will be presented after the rest of the data are analyzed and final corrections for antenna gain and atmospheric attenuation are completed.

D. H. Staelin

B. RADIOMETER AT FOUR MILLIMETER WAVELENGTH

In Quarterly Progress Report No. 66 (page 49) a Millimeter Radio Telescope was described. This report will discuss the development of a radiometer for use with this antenna.

The block diagram of the radiometer is shown in Fig. IV-4. It is a Dicke type of radiometer utilizing a ferrite switch to compare the input signal with a known signal, a single-ended mixer, and two traveling-wave tubes for if amplifiers. The input is

centered at 72 Gc (4.16 mm), the local oscillator at 69.0 Gc, and the if strip has a measured bandpass of approximately 1200 mc. The receiver noise figure was measured to be 31 db, 10 db greater than the design objective. The source

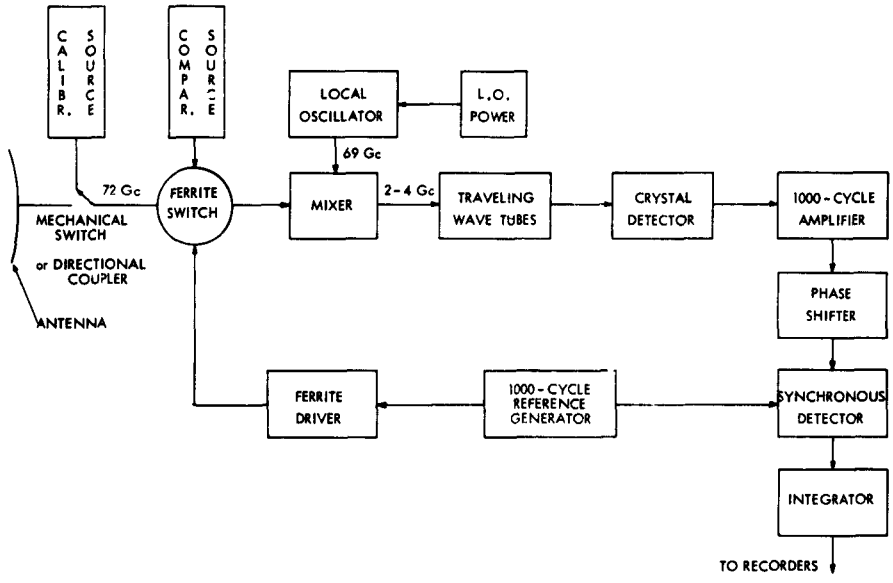


Fig. IV-4. Four millimeter wavelength radiometer.

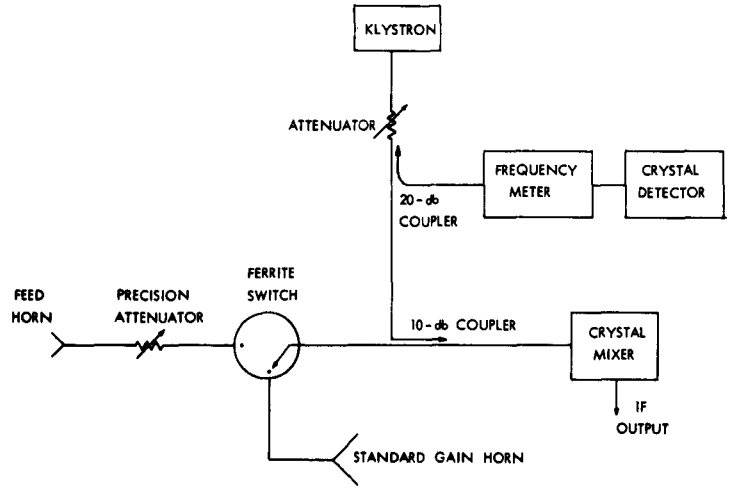


Fig. IV-5. Schematic diagram of millimeter wave section of the radiometer.

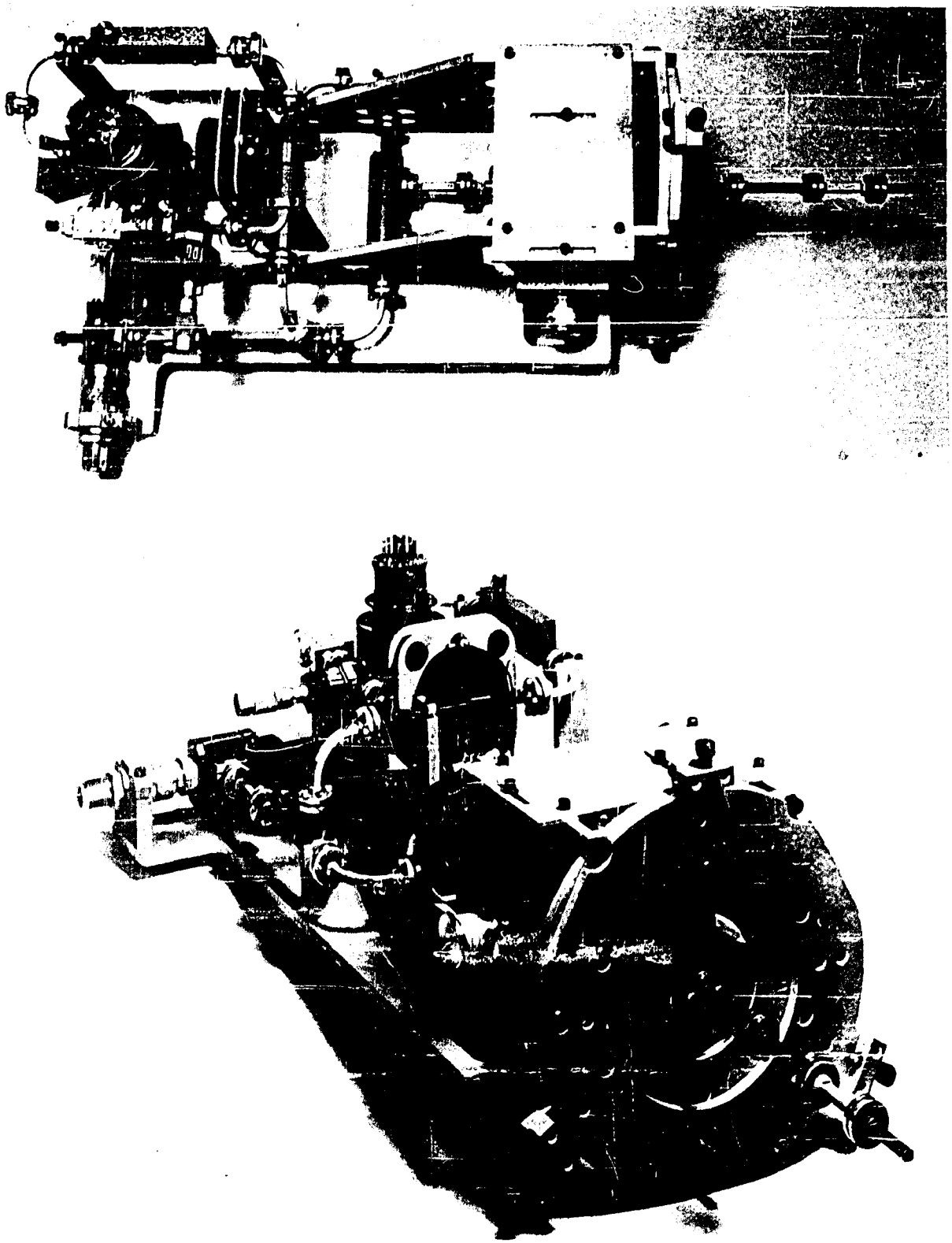


Fig. IV-6. Millimeter wave section of the radiometer.
Upper: top view. Lower: front view.

of difficulty lies in the crystal mixer. An effort is now being made in the laboratory to obtain an improved mixer for use in millimeter-wave radiometers with microwave if strips.

The radiometer was installed in the 10-ft antenna, and an on-axis gain measurement made. The remote transmitter used in testing the 10-ft parabolic antenna was located at Quigley Memorial Hospital at Soldier's Home in Chelsea, a distance of 4.25 miles from the Compton Laboratories Building in Cambridge. The transmitter package consisted of a 6-inch parabolic antenna, klystron, and microwave equipment mounted on an optical telescope. The millimeter-wave section of the radiometer is shown schematically in Fig. IV-5, and pictorially in Fig. IV-6. This section of the radiometer was mounted at the focus of the 10-ft antenna, while the TWT if amplifiers and second detector were mounted in the counterbalance structure at the rear of the antenna.

A standard gain horn is used to provide the comparison signal. When the precision attenuator is tuned to balance, the antenna gain is given by

$$G(db) = G_{cal}(db) + L_{attn}(db) + L_{path}(db)$$

where

G_{cal} = gain of standard gain horn

L_{attn} = calibrated attenuation

L_{path} = difference in path loss.

A comparison technique such as this reduces the effects of variations in transmitter power and atmospheric attenuation.

Weather conditions have not permitted a careful optimization of the feed-horn position. Consequently, the following data indicate only a lower bound on the performance of the antenna. Rough measurements show that the antenna has an on-axis gain of at least 62 ± 2 db, an efficiency of 31 per cent, H-plane (azimuth) side lobes of 12 db and 15 db, and E-plane (elevation) side lobes of 20 db.

J. C. Blinn III

C. OPTICAL RADAR TO STUDY THE EARTH'S ATMOSPHERE

The use of an optical radar to study the Earth's atmosphere by back-scattering has been considered and the construction of an instrument is well advanced.

In particular, we plan to study the dust layers[1] up to heights of 100 km. At first our main object will be some systematic observations of the noctilucent clouds (~80 km); these clouds are more easily distinguishable from molecular Rayleigh scattering, which is predominant at lower heights. The mechanism of Raman scattering will also be

utilized to infer density profiles for the various atmospheric molecular species because the value of the wavelength shift is a characterictic of each species.

The optical radar consists of a short-pulse ruby laser capable of the following performance:

energy in one pulse,	0.5-1.0 joule
length of pulse,	1 μsec
pulse repetition frequency,	1 cps.

The laser beam is collimated by means of a refractor of 2.019-m focal length, and 0.075-m diameter. The receiver utilizes a Newtonian reflecting telescope of 2.70-m focal length and 0.30-m diameter, and a photoelectric photometer.

G. Fiocco

References

1. Following are a few of the numerous references on the subject of cosmic dust and noctilucent clouds: E. J. Öpik, Irish Astron. J. 4, 84 (1956); E. H. Ludlam, Tellus 9, 341 (1957); C. E. Junge and J. E. Manson, J. Geophys. Research 66, 2163 (1961); F. L. Whipple, Nature 189, 127 (1961); G. Witt, Tellus 14, 1 (1962).

V. NOISE IN ELECTRON DEVICES*

Prof. H. A. Haus
Prof. P. L. Penfield, Jr.
Prof. R. P. Rafuse
W. D. Rummler

A. PHOTON STATISTICS OF OPTICAL MASER OUTPUT

The measurement of the output-power spectra of narrow-band optical-maser oscillators or amplifiers is currently under way. Information on the power spectra may be obtained from the statistics of the output photon number. These, in turn, may be inferred from the count of photoelectrons from a photosurface exposed to the optical-maser beam. We shall first present the classical theory of such an experiment, and then summarize the preliminary experimental results.

We assume that the probability of producing a photo-electron current pulse within the time interval dt is $\alpha P(t)\,dt$; P(t) is the power in the incident light (in general, a statistical quantity), and α is a factor incorporating the photoefficiency of the photo-tube. The probability of obtaining exactly K counts in an interval T follows the Poisson law for nonstationary processes.

$$p(K) = \frac{n^K}{K!} e^{-n}, \tag{1}$$

$$n = \int_0^T \alpha P(t)\,dt. \tag{2}$$

If P(t) is a statistical variable, an average has to be taken with respect to its probability distribution. We shall indicate this last average by a bar.

$$\overline{n_T} = \sum_{K=1}^{\infty} p(K)K = \int_0^T \alpha\,\overline{P(t)}\,dt = \alpha T \overline{P} \tag{3}$$

and

$$\overline{n_T^2} = \sum_{K=1}^{\infty} \overline{K^2 p(K)} = \overline{n_T} + \left(\int_0^T \alpha P(t)\,dt\right)^2 . \tag{4}$$

The last integral can be evaluated easily in two limits; for $T \ll \tau_0$, and $T \gg \tau_0$, where τ_0 is the correlation time of P(t) which is of the order of the inverse bandwidth of the maser. (For optical-maser amplifiers with highly reflecting mirrors this time will in general not be shorter than 10 μsec. For an oscillator it will be much longer than

*This work was supported in part by Purchase Order DDL B-00368 with Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology with the joint support of the U.S. Army, Navy, and Air Force under Air Force Contract AF19(604)-7400.

that.) In the limit $T \ll \tau_o$,

$$\overline{\delta n_T^2} = \overline{n_T^2} - \bar{n}_T^2 = a\bar{T}\bar{P} + a^2T^2\overline{P^2(0)} - a^2T^2\bar{P}^2. \tag{5}$$

For a Gaussian amplitude distribution we have[1]

$$\overline{P^2} = 3\bar{P}^2 \tag{6}$$

and thus

$$\overline{\delta n_T^2} = \bar{n}_T(1+2\bar{n}_T). \tag{7}$$

This result applies for a linearly polarized light output. If the light is unpolarized, the mean-square fluctuations are additive, and we have

$$\overline{\delta n_T^2} = \bar{n}_T(1+\bar{n}_T). \tag{8}$$

For a sinusoid, this is

$$\overline{\delta n_T^2} = \bar{n}_T. \tag{9}$$

Thus, for a photoelectron count $\bar{n}_T \gg 1$ within a time $T \ll \tau_o$, there is a great difference between the fluctuations observed for a Gaussian signal and that of a sinusoid (or near sinusoid).

In the limit $T \gg \tau_o$, we may approximate

$$\left[\int_0^T P(t)\,dt\right]^2 \cong T\int_0^T \overline{P(t)\,P(t+\tau)}\,dt. \tag{10}$$

For a Gaussian process[2]

$$\overline{P(t)\,P(t+\tau)} = \bar{P}^2 + 2R^2(\tau), \tag{11}$$

where $R^2(\tau)$ is the square of the amplitude autocorrelation function that is related to the spectrum, and thus to the line shape, by a Fourier transform. We obtain for linearly polarized light (compare Purcell[3])

$$\overline{\delta n_T^2} \cong \bar{n}_T\left(1 + \bar{n}_T\,\frac{\tau_o}{T}\right). \tag{12}$$

Here, we have set

$$\tau_o = \frac{2\int_{-\infty}^{+\infty} R^2(\tau)\,d\tau}{\bar{P}^2}. \tag{13}$$

The limiting value of $\overline{\delta n_T^2}/\bar{n}_T$ attained for large T, if the photon flux is held constant (so that $\bar{n}_T$ is proportional to T) gives the correlation time τ_o.

We have made preliminary measurements on an optical maser oscillator of 6328 Å center frequency and 1-mw output, using a phototube (EMI/9558) and a high-speed counter with a maximum counting rate of 10 counts/μsec. The light was attenuated before reaching the multiplier and the low spikes were rejected in the count. The shortest counting interval used was 100 μsec, with $\bar{n}_T = 5.5$. If the maser signal had been Gaussian, we should have seen the large $\overline{\delta n_T^2}$ corresponding to Eqs. 7 or 8. This was not observed. We found $\overline{\delta n_T^2} = \beta \bar{n}_T$, for $T = 10^{-4}$ to 10 sec in decade steps, where β was a number varying in magnitude between 0.5 and 12, the large deviation from unity occurring for relatively long observation times T and large average photon counts. The deviations of β from unity may be due in part to the statistics of the photomultiplication. We believe that the experiment has shown that the signal is not Gaussian, exhibiting power fluctuations much smaller than those corresponding to a Gaussian signal of a bandwidth that one may reasonably expect for a maser oscillator. The experiments are continuing. The experiments are undertaken jointly with C. Freed and R. J. Carbone of Lincoln Laboratory, M.I.T., and J. McDonald, a senior in the Department of Electrical Engineering, M.I.T.

G. Fiocco, H. A. Haus

References

1. W. B. Davenport, Jr. and W. L. Root, Random Signals and Noise (McGraw-Hill Book Company, New York, 1958), p. 254.

2. Ibid, p. 255.

3. E. M. Purcell, The question of correlation between photons in coherent light rays, Nature 178, 1447 (December 29, 1956).

PLASMA DYNAMICS

VI. PLASMA PHYSICS*

Prof. S. C. Brown
Prof. G. Bekefi
Prof. D. R. Whitehouse
M. L. Andrews
V. Arunasalam
J. F. Clarke
J. D. Coccoli
F. X. Crist
H. Fields
E. W. Fitzgerald, Jr.
W. H. Glenn, Jr.
E. B. Hooper, Jr.
J. C. Ingraham
P. W. Jameson
R. L. Kronquist
D. T. Llewellyn-Jones
J. J. McCarthy
W. J. Mulligan
J. J. Nolan, Jr.
H. R. Radoski
G. L. Rogoff
F. Y-F. Tse
R. E. Whitney

A. INTERACTION BETWEEN THE RADIATION FIELD AND THE STATISTICAL STATE OF A PLASMA

The fact that the electromagnetic radiation generated by a plasma can alter the statistical distribution of the radiating particles is generally not taken into consideration. The effects are expected to be large when, for example, the plasma departs from thermodynamic equilibrium and becomes unstable. The intense radiation that the plasma can now generate[1] reacts on the energy distribution of the particles. And it is not inconceivable that by driving the distribution closer to a Maxwellian, the instability will in time be quenched.

The problem of coupling between the radiation field and the statistical state of the plasma has been discussed recently by Kudryavtsev,[2] Akhiezer et al.,[3] and Dreicer.[4] A self-consistent calculation requires a simultaneous solution of two nonlinear equations and thus far has not been carried out. In this report we make no attempt to solve the problem – we shall merely set up the appropriate equations and present them in a form different from the one given by the above-mentioned authors. Since much of the intense radiation occurs at radio and microwave frequencies and the emission is primarily caused by the free electrons, we shall restrict our considerations to this situation.

We have two equations that describe a system composed of a plasma and its radiation field. The one is the equation of transfer for the flow of radiant energy. If I_ω is the intensity of radiation in the radian frequency interval between ω and $\omega + d\omega$ along a ray $\vec{s}$, then

$$\frac{1}{c}\frac{\partial I_\omega}{\partial t} + \frac{\partial I_\omega}{\partial s} = j_\omega - \alpha_\omega I_\omega \tag{1}$$

where j_ω and α_ω are the emission and absorption coefficients of the plasma, respectively. The foregoing equation can be rewritten[1] in terms of the rate of spontaneous emission of

*This work was supported in part by the U.S. Atomic Energy Commission (Contract AT(30-1)-1842); and in part by the U.S. Air Force (Electronic Systems Division) under Contract AF19(604)-5992.

radiation, η_ω, and the distribution of electron velocities, f(v):

$$\frac{1}{c}\frac{\partial I_\omega}{\partial t} + \frac{\partial I_\omega}{\partial s} = \int \eta_\omega f(v)\, 4\pi v^2 dv + \left[\frac{8\pi^3 c^2}{\omega^2}\int \eta_\omega(v)\,\frac{\partial f(v)}{\partial u}\,4\pi v^2\, dv\right] I_\omega. \tag{2}$$

Here $u = mv^2/2$ is the electron energy, and f(v) has been assumed to be isotropic (in principle there is no difficulty in extending Eq. 2 to anisotropic distributions).

The second equation is Boltzmann's equation that prescribes the distribution of electron velocities and concentrations in the presence of fields and particle interactions:

$$\frac{\partial f}{\partial t} + \underset{\sim}{v}\cdot\nabla_{\underset{\sim}{r}} f - \frac{e}{m}\left[\underset{\sim}{E} + \underset{\sim}{v}\times\underset{\sim}{B}\right]\cdot\nabla_v f = C_{\text{collision}} + C_{\text{radiation}}. \tag{3}$$

$\vec{E}$ and $\vec{B}$ are the externally applied fields. Since, however, I_ω can be considered as the electromagnetic radiation composed of externally applied fields, as well as the internally generated radiation, $\vec{E}$, $\vec{H}$ represent only that portion of the external forces not included in I_ω.

The term $C_{\text{radiation}}$ (hereafter C_{rad}) is a way of writing symbolically the difference between the rates at which electrons are scattered into and out of a volume element of velocity space as a result of their emission and absorption of radiation; it is this term that is generally neglected in computations of f and which, as we shall see, makes Eq. 3 a function of I_ω.

We shall now calculate C_{rad}. The quantity $C_{\text{rad}}d^3v$ equals the rate at which electrons enter the volume element d^3v minus the rate at which they leave d^3v, as a result of spontaneous emission, absorption, and stimulated emission. We shall consider, first, the rates at which they enter and leave d^3v as a result of radiative transitions to and from two neighboring elements $d^3(v')$ and $d^3(v'')$, where $v' < v$ and $v'' > v$. The speeds v, v', and v'' are such that transitions $v \rightleftarrows v'$, $v \rightleftarrows v''$ are associated with the emission or absorption of a photon $\hbar\omega$ whose frequency is given by

$$\begin{aligned}\hbar\omega &= (1/2)\, mv^2 - (1/2)\, mv'^2 \\ &= (1/2)\, mv''^2 - (1/2)\, mv^2.\end{aligned} \tag{4}$$

Writing $\eta_{\omega A}$ as the rate of absorption and $\eta_{\omega S}$ as the rate of stimulated emission, we obtain

$$\begin{aligned}C_{\text{rad}}(v,v',v'')\, d^3v = &\left[\eta_{\omega A}(v')\, f(v')\, d^3v' - \eta_{\omega A}(v)\, f(v)\, d^3v\right]\frac{I_\omega}{\hbar\omega}\, d\omega\, d\Omega \\ &+ \left[\eta_{\omega S}(v'')\, f(v'')\, d^3v'' - \eta_{\omega S}(v)\, f(v)\, d^3v\right]\frac{I_\omega}{\hbar\omega}\, d\omega\, d\Omega \\ &+ \left[\eta_\omega(v'')\, f(v'')\, d^3v'' - \eta_\omega(v)\, f(v)\, d^3v\right]\frac{d\omega\, d\Omega}{\hbar\omega}\end{aligned} \tag{5}$$

where $\eta_{\omega A}$, $\eta_{\omega S}$, and η_ω are related as in Einstein's A and B coefficients:

$$\left.\begin{aligned} \eta_\omega(v') &= \frac{\hbar\omega^3}{8\pi^3 c^2}\,\eta_{\omega S}(v') \\ v\eta_{\omega A}(v) &= v'\eta_{\omega S}(v') \end{aligned}\right\} \tag{6}$$

We simplify Eq. 5 by proceeding to the low-frequency limit $\hbar\omega \to 0$. By means of Eqs. 6 we first replace $\eta_{\omega A}$ and $\eta_{\omega S}$ by the rate of spontaneous emission, η_ω. We then expand $f(v')$, $f(v'')$, $\eta_\omega(v')$, and $\eta_\omega(v'')$ in Taylor series about the velocity v, retaining three terms of the expansion wherever I_ω appears as a multiplier in Eq. 5, and two terms when I_ω does not enter. After somewhat lengthy algebraic manipulations and use of Eqs. 4 we find that

$$C_{rad}(v,v',v'') = \frac{1}{mv^2}\left[\frac{d[v\eta_\omega(v)\,f(v)]}{dv} + \frac{8\pi^3 c^2}{\omega^2}\,I_\omega\,\frac{d}{mdv}\left[\eta_\omega(v)\,\frac{df(v)}{dv}\right]\right] d\omega\, d\Omega. \tag{7}$$

The total value of C_{rad} is obtained by integrating Eq. 7 over all frequencies ω of the low-frequency spectrum and over all angles $d\Omega$ in which radiation flows. Since I_ω and η_ω refer to one characteristic wave only, we must sum over both modes of polarization. Thus,

$$C_{rad} = \sum_{1,2}\int_\omega\int_\Omega \frac{1}{mv^2}\left[\frac{d(v\eta_\omega f)}{dv} + \frac{8\pi^3 c^2}{\omega^2}\,I_\omega\,\frac{d}{mdv}\,\eta_\omega\,\frac{df}{dv}\right] d\omega\, d\Omega. \tag{8}$$

As a check on the algebraic manipulations, we multiply Eq. 8 by $4\pi v^2\,dv$ and integrate over all velocities. We find the correct result, $\int_0^\infty C_{rad}\,4\pi v^2\,dv = 0$, which shows that electrons are conserved. Knowing the form of C_{rad}, we can in principle evaluate I_ω and f from Eqs. 2 and 3 for any radiation mechanism, η_ω, of a plasma subjected to external fields $\vec{E}$, $\vec{B}$. Note that all of our results apply in the nonrelativistic limit. Extension of the results to plasmas with relativistic electrons is straightforward.

The two terms on the right-hand side of Eq. 8 refer to different physical processes. The first term describes the net rate at which particles leave a certain energy range as a result of energy loss by spontaneous emission. The second term represents the rate of "heating" of electrons by the ambient radiation field, I_ω.

Special Cases

(a) Consider a collisionless plasma which is in a steady state in the presence of the radiation field I_ω, which in part is due to the radiation from the plasma, and in part to any external sources that may be present. Then, from Eq. 3,

$$\frac{\partial f}{\partial t} = C_{rad} = 0. \tag{9}$$

Evaluating Eqs. 8 and 9, we obtain

$$f = \text{constant} \times \exp\left[- \int_0^v \frac{\Sigma\int\int \eta_\omega}{\Sigma\int\int \frac{8\pi^3 c^2}{\omega^2} I_\omega \eta_\omega} \, mv \, dv \right], \tag{10}$$

where the symbols $\Sigma\int\int$ denote summation over two polarizations and integrations over ω and Ω.

(b) Suppose that in addition to C_{rad}, the electrons make elastic collisions with atoms or ions. If $\nu(v)$ is the collision frequency,[5] then

$$C_{coll} = G \frac{1}{2v^2} \frac{d}{dv}\left[v^2 \nu(v) \left\{ vf + \frac{kT}{m} \frac{df}{dv} \right\} \right], \tag{11}$$

where $G = 2m/M$ with M as the mass of the atom or ion, and T as their temperature. In the steady state (with $E = B = 0$ as in case (a)),

$$\frac{\partial f}{\partial t} = C_{rad} + C_{coll} = 0. \tag{12}$$

Solving, we obtain

$$f \propto \exp\left[- \int_0^v \frac{\left[1 + \frac{\Sigma\int\int \eta_\omega}{uG\nu} \right] mv \, dv}{kT_g + \Sigma\int\int \frac{8\pi^3 c^2}{\omega^2} I_\omega \frac{\eta_\omega}{uG\nu}} \right]. \tag{13}$$

G. Bekefi

References

1. G. Bekefi, J. L. Hirshfield, and S. C. Brown, Phys. Fluids 4, 173 (1961); Phys. Rev. 122, 1037 (1961).

2. V. S. Kudryavtsev, The electron distribution function in a plasma in a magnetic field, Plasma Physics and the Problem of Controlled Thermonuclear Reactions, Vol. III, edited by M. A. Leontovich (Pergamon Press, London and New York, English edition, 1959), pp. 133-140.

3. A. I. Akheizer, V. F. Aleksin, V. G. Bar'Yakhathr, and S. V. Peletminskii, Soviet Phys.—JETP 15, 386 (1962).

4. H. Dreicer, Bull. Am. Phys. Soc. 7, 632 (1962).

5. W. P. Allis, Motions of Ions and Electrons, Technical Report 299, Research Laboratory of Electronics, M. I. T., June 13, 1956.

B. ELECTRON TEMPERATURE DECAY IN THE AFTERGLOW OF A PULSED HELIUM DISCHARGE

The Transient Microwave Radiation Pyrometer[1] has been used to study the electron

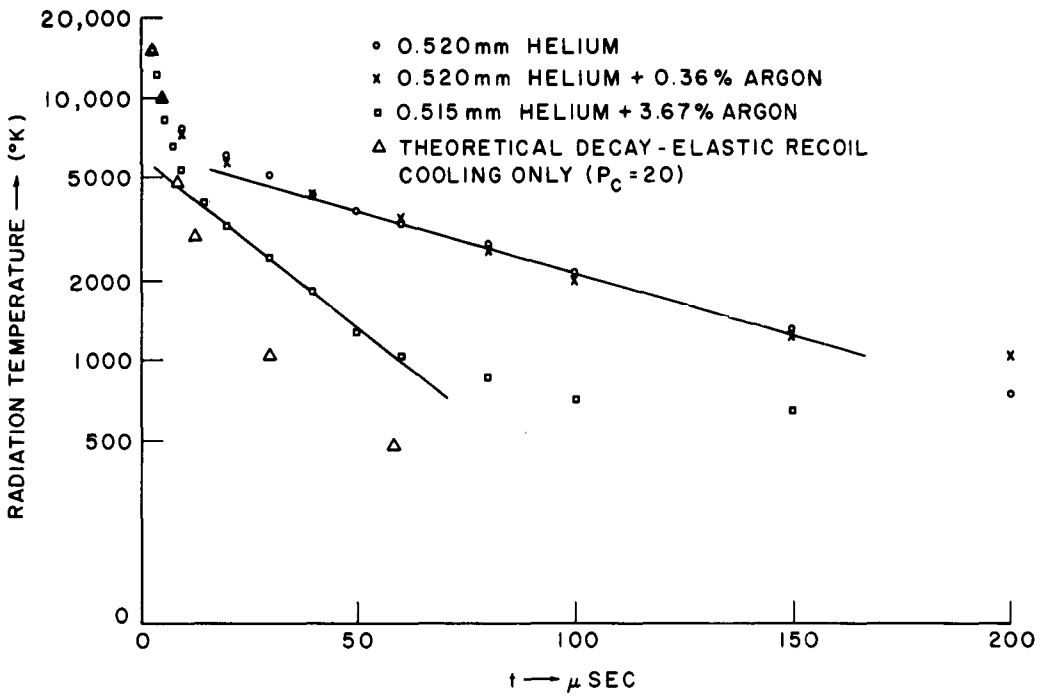


Fig. VI-1. Electron temperature decay for varying amounts of added argon impurity.

temperature decay in the helium afterglow.[2]

Figure VI-1 illustrates the dependence of the temperature decay on fractional additions of argon of zero per cent, 0.36 per cent, and 3.67 per cent at a helium pressure of 0.520 mm Hg. As we have noted, the asymptotic temperature decay rate at this pressure should be equal to the rate of destruction of metastable atoms.[2] The increase in this asymptotic decay rate with increasing argon concentration bears out our hypothesis that metastable atoms are determining the asymptotic behavior of the temperature decay. From the observed asymptotic decay rates in Fig. VI-1, a cross section for the argon-metastable destructive collision can be calculated. A value of 2.6×10^{-16} cm^2 is obtained which compares reasonably with Biondi's[3] measurement of the argon ionization cross section in an argon-metastable collision of 9×10^{-17} cm^2.

Figure VI-2 shows the temperature decay for lower pressures at which, presumably, the metastables are lost by diffusion at such a rapid rate that their heating effect is not felt in the afterglow. The electron temperature decay in this case should now be determined by elastic recoil energy losses. The theoretical curve for such a decay may be readily calculated for helium for which the collision probability is known fairly well to be $P_c = 20p$ cm^{-1}, where p is in mm Hg. The pressure multiplied by the time is the

normalized time variable, so that the curves should superpose if only elastic recoil is operating to cool the electrons. The curve at p = 0.120 mm Hg displays some residual

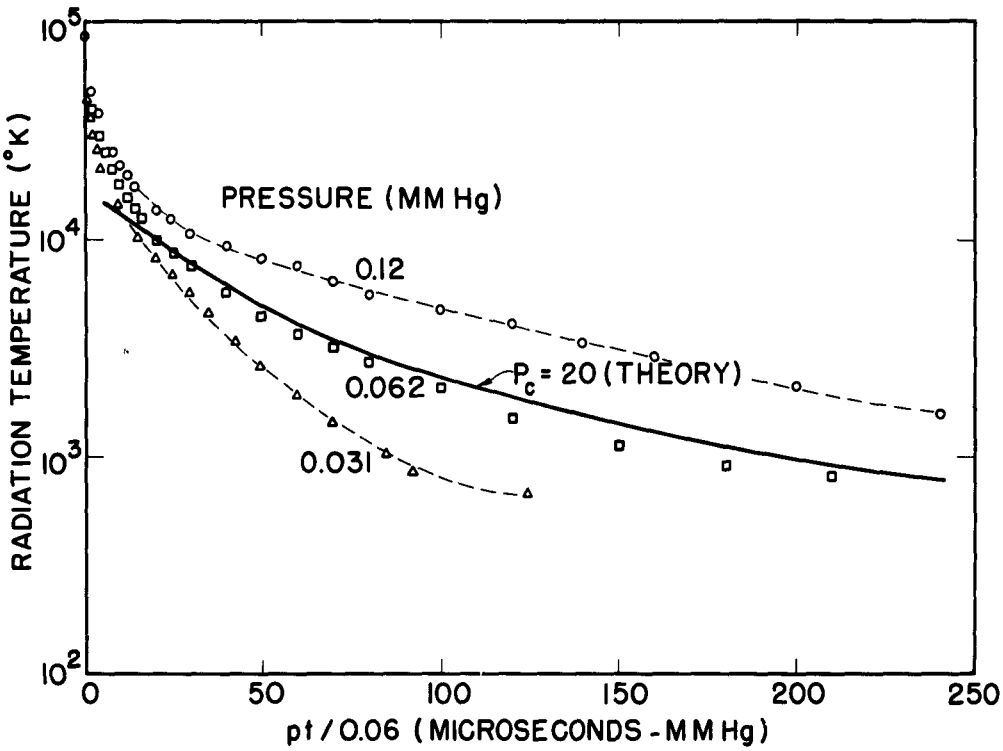


Fig. VI-2. Electron temperature decay for low pressures.

metastable heating, while the curve at p = 0.03 mm Hg displays some apparent additional cooling. Calculations have indicated that diffusion cooling should not contribute significantly in this region.

J. C. Ingraham

References

1. J. C. Ingraham and J. J. McCarthy, A transient microwave radiation pyrometer, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M.I.T., January 15, 1962, pp. 76-79.

2. J. C. Ingraham, Electron temperature decay in a helium plasma afterglow, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 83-86.

3. M. A. Biondi, Phys. Rev. 83, 653 (1951).

C. MEASUREMENTS OF ELECTRON RELAXATION RATES IN PLASMAS

The thermodynamic state of a plasma can be inferred from its radiation spectrum. In this report we present measurements of the radiation spectrum in the afterglow of a plasma. By observing the approach of the spectrum toward one associated with the equilibrium state, it is possible to infer the relaxation rates to the equilibrium condition. We shall show that in our plasma two relaxation rates are dominant. At high pressures the relaxation toward a Maxwellian distribution is governed by short-range electron-atom collisions, while at low pressures the dominant relaxation mechanism comes from electron-electron impacts.

In previous reports[1,2] we have shown that the microwave radiation spectrum from weakly ionized plasmas departs significantly from the Kirchhoff-Planck law. The departure has been shown to be consistent with theory, provided that one assumed a non-Maxwellian energy distribution for the plasma electrons. The magnitude of the departure also depends on the energy dependence of the electron-neutral collision frequency for momentum transfer in the energy range where most of the electrons are situated. By comparing the measured and calculated spectra, it has been possible to deduce the electron energy distribution function and the mean energy.

Recent measurements further substantiate the above-mentioned results. We measure the radiation temperature T_r (as defined previously[1,2]) of a positive column in argon, immersed in a dc magnetic field as a function of time in the afterglow of the plasma. A discharge of 1-msec width is pulsed at a frequency of 200 cps, and the radiation temperature is measured as a function of magnetic field ($B = \omega_b \frac{m}{e}$, where ω_b is the electron cyclotron frequency), at a fixed time in the afterglow. The measurements were performed with the transient radiation pyrometer described in a previous report,[3] and the measuring frequency was fixed at $6\pi \times 10^9$ rad/sec.

The pyrometer accepts radiation within a 1-μsec gate, which is movable in time from 10 μsec before the discharge pulse is shut off to 3 msec past this cutoff point. We take this point to be our reference in time (i.e., t=0). so that when t is negative we refer to results obtained from a "going" discharge, and when t is positive we refer to the results obtained in the afterglow.

Some of the results are shown in Fig. VI-3. In all cases the radiation temperature is shown as a function of frequency or magnetic field for various times in the afterglow; p_o is the pressure reduced to 0°C, and I is the discharge current. Figure VI-3a and VI-3b gives the results at a low pressure for two discharge currents. Figure VI-3c and VI-3d shows the results for the same currents but at a higher pressure. The departure of the radiation temperature from the Kirchhoff-Planck law (i.e., the peaks at cyclotron resonance, $\omega = \omega_b$) is seen to be decreasing as a function of time in the afterglow. This is consistent with the interpretation that these peaks are due to non-Maxwellian

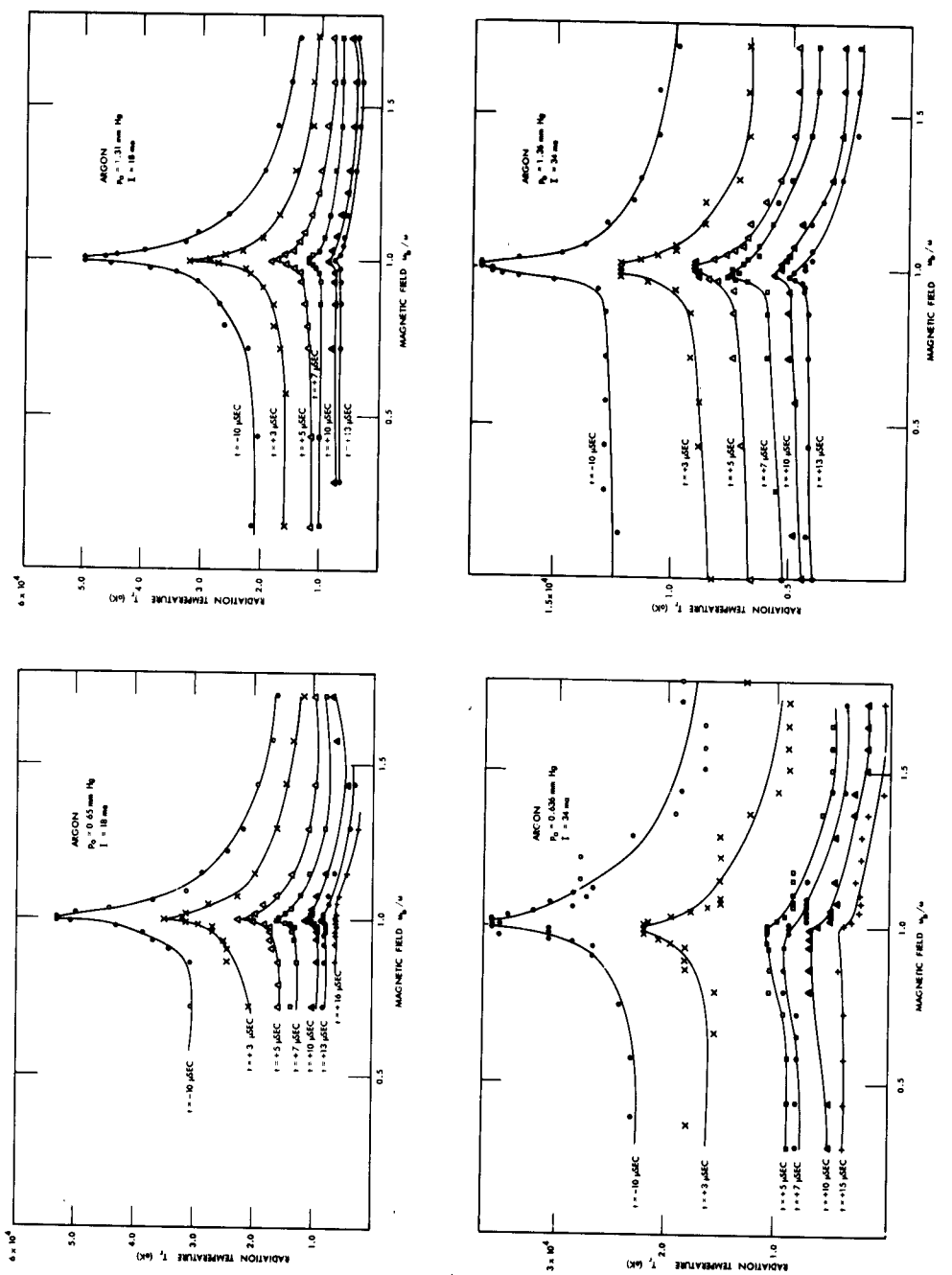


Fig. VI-3. Radiation temperature as a function of magnetic field for various times in the afterglow.

electron velocity distributions. As the discharge pulse is cut off, the electrons relax to a Maxwellian distribution and the peaks disappear. The approach to a Maxwellian distribution may be due to electron-electron collisions or electron-neutral collisions. At low pressures it is the electron-electron collisions that dominate the relaxation process. Here we would expect the rate of relaxation to be strongly dependent on the electron density.[4] This is borne out by the results shown in Fig. VI-3a and VI-3b, in which the peak in T_r is seen to decrease much faster for a discharge current of 34 ma than for a current of 18 ma. Figure VI-3c and VI-3d shows that the relaxation rate (as interpreted by the decrease of the peaks in T_r) at a higher pressure, for which electron-neutral collisions dominate the process, is relatively insensitive to discharge current.

The results can only be interpreted qualitatively, at this time, for two reasons. First, the radiation temperature does not relate linearly to the distribution function, and second, we did not measure the electron density in the afterglow. We are now conducting calculations that will enable us to determine the distribution function and the mean energy of the electrons as a function of time in the afterglow by the methods previously described.[1,2] We have installed a probe in our discharge tube in order to measure the electron density in the afterglow. With these results we hope to determine a time constant for the approach of the electrons to a Maxwellian distribution.

H. Fields, G. Bekefi

References

1. H. Fields and G. Bekefi, Determination of electron energies and energy distributions from measurements of the nonthermal radiation from plasmas, Quarterly Progress Report No. 65, Research Laboratory of Electronics, M.I.T., April 15, 1962, pp. 73-79.

2. H. Fields, G. Bekefi, and S. C. Brown, Microwave emission from non-Maxwellian plasmas, Phys. Rev. 129, 506 (1963).

3. J. C. Ingraham and J. J. McCarthy, A transient microwave radiation pyrometer, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M.I.T., January 15, 1962, pp. 76-79.

4. L. Spitzer, Jr., Physics of Fully Ionized Gases (Interscience Publishers, Inc., New York, 1956), p. 78.

D. VISCOUS DAMPING OF PLASMA WAVES

An attempt will be made to explain the experimental results on propagation and damping of ion acoustic waves reported by Wong, D'Angelo, and Motley.[1] The momentum transport equations will be modified by the introduction of a viscous force. This procedure was suggested by a comparison of the wavelength and Coulomb mean-free path. For a frequency of 10^5 cps and a phase velocity of 10^5 cm/sec, the associated wavelength is 1 cm. For a particle density of $6 \times 10^{10}/cm^3$ and a temperature of

2300°K, the Debye length squared is approximately 1.8×10^{-6} cm^2. Hence, taking the magnitude of the mean-free path as $L \approx \left(n\lambda_D^3\right)\lambda_D$, where n is the density, and λ_D is the Debye length, we obtain $L \approx 0.2$ cm. Because of the similarity of the magnitudes of wavelength and mean-free path, we suspect that the waves will be strongly viscous-damped. For the magnitude of the viscosity coefficient we take $\eta \approx LV_T\rho$, where V_T is the thermal velocity, and ρ is the mass density. Hence the kinematic viscosity will be $\nu = \eta/\rho \approx LV_T$.

The equations that will be used to obtain the dispersion relation are the linearized force and continuity equations and Poisson's equation.

$$\frac{\partial \vec{\Gamma}_\pm}{\partial t} = -\mu_{s_\pm}^2 \nabla n_\pm \pm \frac{en_o}{m_\pm}\vec{E} + \nu_\pm \nabla^2 \vec{\Gamma}_\pm \tag{1}$$

$$\frac{\partial n_\pm}{\partial t} + \nabla \cdot \bar{\Gamma}_\pm = 0 \tag{2}$$

$$\nabla \cdot \bar{E} = 4\pi e(n_+ - n_-). \tag{3}$$

In these equations the particle flux $\vec{\Gamma}_\pm$ is defined as $n_o\vec{V}_\pm$, where $\vec{V}_\pm$ is the ion or electron velocity, and n_o is the unperturbed particle density; $n_\pm$ is the perturbed particle density; and $\mu_{s_\pm}$ is the ion or electron sound velocity. For equal electron and ion temperatures we have $\mu_{s_-}^2 = m_+/m_-\mu_{s_+}^2$ and $\nu_-^2 = m_+/m_-\nu_+^2$. By using Poisson's equation and the continuity equations, the force equations may be written in terms of the perturbed densities.

$$\frac{\partial^2 n_\pm}{\partial t^2} = \mu_{s_\pm}^2 \nabla^2 n_\pm \mp \omega_{p_\pm}^2 (n_+ - n_-) + \nu_\pm \nabla^2 \frac{\partial n_+}{\partial t}, \tag{4}$$

where ω_p is the plasma frequency, and

$$\omega_{p_\pm}^2 = \frac{4\pi n_o e^2}{m_\pm}.$$

By allowing $n_\pm$ to vary as $e^{j(\bar{k}\cdot\bar{r} - \omega t)}$ in Eq. 4, the following equations are obtained:

$$\begin{aligned}\left[\omega^2 - \mu_{s_+}^2 k^2 - \omega_{p_+}^2 + j\nu_+\omega k^2\right] n_+ + \omega_{p_+}^2 n_- &= 0 \\ \omega_{p_-}^2 n_+ + \left[\omega^2 - \mu_{s_-}^2 k^2 - \omega_{p_-}^2 + j\nu_-\omega k^2\right] n_- &= 0.\end{aligned} \tag{5}$$

Setting the determinant of Eq. 5 equal to zero, the following dispersion relation results:

$$\frac{\omega_{p_+}^2}{\omega^2 - k^2\left(\mu_{s_+}^2 - j\nu_+\omega\right)} + \frac{\omega_{p_-}^2}{\omega^2 - k^2\left(\mu_{s_-}^2 - j\nu_-\omega\right)} = 1. \tag{6}$$

In the experiment of Wong and others,[1] ω^2 varies from $\sim 4 \times 10^7$ to 4×10^{11}. For $n_o = 6 \times 10^{10}/cm^3$, $\omega_{p_-}^2$ will be approximately 1.8×10^{20}, and $\omega_{p_+}^2$, for $m_+/m_- = 10^5$, will be 1.8×10^{15}. By using this information, as well as the order of magnitude of the sound velocity and the kinematic viscosity, an approximate equation for the ion waves can be written from Eq. 6.

$$k^2 = \frac{\omega^2}{2\mu_{s_+}^2 - j\omega\nu_+}. \tag{7}$$

Setting $k = k_o + i\mathcal{K}$, where k_o indicates the propagation, and $\mathcal{K}$ the damping, Eq. 7 yields

$$k_o = \frac{\omega}{2\mu_{s_+}} X(t)$$

$$\mathcal{K} = \frac{\omega}{2\mu_{s_+}} Y(t), \tag{8}$$

where

$$t = \left(\frac{\omega\nu_+}{2\mu_{s_+}^2}\right)$$

and

$$X(t) = \left[\frac{(1+t^2)^{1/2} + 1}{1 + t^2}\right]^{1/2}$$

$$Y(t) = \left[\frac{(1+t^2)^{1/2} - 1}{1 + t^2}\right]^{1/2}.$$

In Fig. VI-4 we show the quantities $\left(\nu_+/\mu_{s_+}\right)\mathcal{K}$ and $\left(1/2\mu_{s_+}\right)\omega/k = V_{phase}/2\mu_{s_+}$ as functions of $\left(\nu_+\omega/2\mu_{s_+}^2\right)$. It should be mentioned that the phase velocity, from simple Navier-Stokes theory, rises much more rapidly with increasing ω than higher order theory or experiments indicate, while the damping remains adequately described by the simple theory.[2,3]

In Fig. VI-5 we present the first attempt to fit the damping theory to the experimental

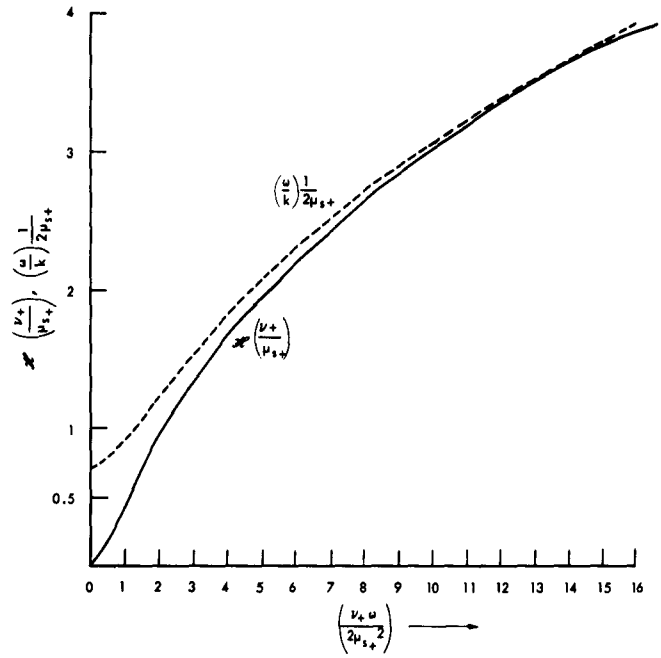


Fig. VI-4. Theoretical propagation and damping functions.

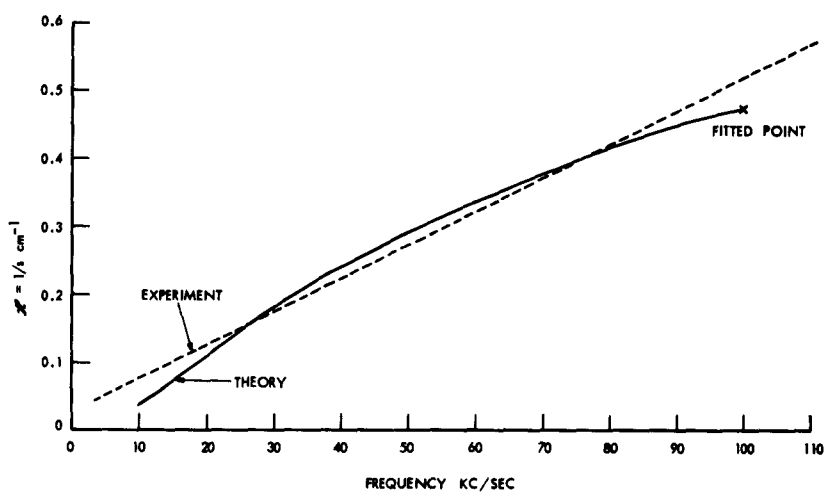


Fig. VI-5. $\mathcal{K}$ as a function of frequency.

data (Fig. 2 of Wong et al.[1]). The data give $\mathcal{K} \approx 0.47$ for potassium at a frequency of 10^5 cps. At this frequency we choose $\left(\nu_+ \omega / 2\mu_{s_+}^2\right) = 5$. Since, from Fig. VI-4, $\nu_+ / \mu_{s_+} \mathcal{K} = 1.96$ at this point, we have $\mu_{s_+} / \nu_+ = 0.24$. This guess is seen to give a fairly good representation of the experimentally observed damping.

The phase velocity obtained from our fitted data is high because for μ_{s_+} itself we obtain 2.6×10^5 cm/sec, which is approximately the observed phase velocity.[1] Our computed phase velocity will vary from 4×10^5 cm/sec at 10 kc/sec almost linearly to 11×10^5 cm/sec at 100 kc/sec. The fitted point also gives a value of $\nu_+ \approx 11 \times 10^5$ cm^2/sec, which is a correct order of magnitude.

Wong, D'Angelo, and Motley, in their Figure 1, give the experimentally observed phase velocity as a function of frequency.[1] These data appear to indicate slightly higher phase velocities for higher frequencies, as well as higher phase velocities for lower densities. Figure VI-4 shows an increasing phase velocity with frequency and also, if the frequency is held fixed, an increasing phase velocity with decreasing density, since $\nu_+ \sim n\lambda_D^4 \sim 1/n$.

For the sake of completeness, we shall compute the viscosity from Boltzmann theory,

$$\nu_+ = \frac{4}{3} \frac{\tau k T}{m} . \tag{9}$$

In Eq. 9 T is the temperature, k is Boltzmann's constant, m is the ion mass, and τ is the time constant given by Spitzer,[5]

$$\tau = \frac{m^{1/2} (3kT)^{3/2}}{8 \times .714 \pi n e^4 \ln \Lambda} , \tag{10}$$

where $\Lambda = 12\pi^{3/2} n\lambda_D^3$.

For potassium, with $n = 6 \times 10^{10}$/cm^3 and T = 2300°K, we obtain

$$\nu_+ = 0.883 \times 10^5 \text{ cm}^2/\text{sec}. \tag{11}$$

Defining $\sqrt{2}\, \mu_{s_+}$ as 2.5×10^5 cm/sec, the measured value of the phase velocity, we obtain at 100 kc/sec,

$$\frac{\nu_+ \omega}{\left(\sqrt{2}\, \mu_s\right)^2} = 0.886. \tag{12}$$

The theoretical curve and the experimental points for $\mathcal{K}$, the damping constant, as a function of frequency are shown in Fig. VI-6. It is clear that the viscous theory gives the correct order of magnitude of the damping.

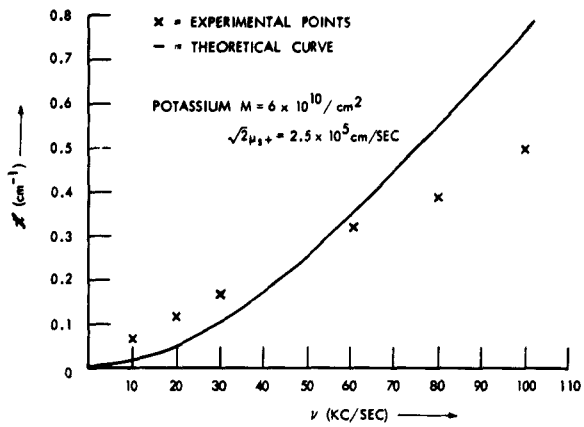


Fig. VI-6. Damping constant as a function of frequency.

Therefore, at present, it seems that a simple viscous theory can explain the observations on ion waves without recourse to Landau damping as discussed by Fried and Gould.[5]

H. R. Radoski

References

1. A. Y. Wong, N. D'Angelo, and R. W. Motley, Propagation and damping of ion acoustic waves in highly ionized plasmas, Phys. Rev. Letters 9, 415-416 (1962).

2. M. Greenspan, Propagation of sound in rarefied helium, J. Acoust. Soc. Am. 22, 568-571 (1950).

3. M. Greenspan, Propagation of sound in five monatomic gases, J. Acoust. Soc. Am. 28, 644-648 (1956).

4. L. Spitzer, Jr., Physics of Fully Ionized Gases (Interscience Publishers, Inc., New York, 1956), p. 78.

5. B. D. Fried and R. W. Gould, Longitudinal ion oscillations in a hot plasma, Phys. Fluids 4, 139-147 (1961).

VII. PLASMA ELECTRONICS*

Prof. L. D. Smullin
Prof. H. A. Haus
Prof. A. Bers
Prof. W. D. Getty
Prof. D. J. Rose
Prof. T. H. Dupree
Prof. L. M. Lidsky
Prof. E. P. Gyftopoulos
Dr. G. Fiocco
F. Alvarez de Toledo
W. L. Brassert
R. J. Briggs
J. R. Cogdell
L. J. Donadieu
S. A. Evans
E. T. Gerry
B. A. Hartenbaum
H. Y. Hsieh
G. I. Kachen
P. K. Karvellas
J. D. Levine
L. N. Lontai
D. L. Morse
R. T. Nowak
L. M. Petrie, Jr.
W. D. Rummler
A. J. Schneider
P. E. Serafim
P. S. Spangler
G. Theodoridis
E. Thompson
C. E. Wagner
S. Wilensky
H. L. Witting
J. C. Woo

A. BEAM-PLASMA DISCHARGE

When a strong electron beam is sent through a partially evacuated region, a series of events occurs that have led us to coin the name "beam-plasma discharge." By this we mean a dynamic process in which plasma is produced by electrons that have been accelerated in microwave fields, and the microwaves in turn have been generated by the strong interaction between the beam and the plasma.[1] The interaction between the beam and plasma is such that intense microwave oscillations are produced either in the neighborhood of the electron cyclotron frequency or of the electron plasma frequency, ω_p. The oscillations are accompanied by large oscillatory velocities of the plasma electrons. When these velocities become large enough, the plasma electrons will further ionize the ambient gas, thus shifting ω_p and the oscillation frequency.

Simultaneous with the ionization process, considerable rf emission and light emission occur and large ion currents appear at the walls of the cylindrical waveguide. Some of the plasma electrons are accelerated to very high energies by the fields in the discharge. This is inferred from the x-rays emitted by the walls and by a molybdenum target, as well as from the large negative floating potential of a probe placed near the beam.

1. Plasma Electron Energy

Four different types of measurements have been used to determine the electron energy in the plasma.

(a) An average electron temperature has been calculated from measurements of ion currents leaving the plasma volume, and of the loss of power by the electron beam. In the steady state, the equation of continuity is

*This work was supported in part by the National Science Foundation (Grant G-24073); in part by the U.S. Navy (Office of Naval Research) under Contract Nonr-1841(78); and in part by Purchase Order DDL B-00368 with Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology with the joint support of the U. S. Army, Navy, and Air Force under Air Force Contract AF 19(604)-7400.

$$I_+ = I_- = n_- e \nu_i V_p, \tag{1}$$

where I_+ is the total ion current, I_- is the total electron current leaving the plasma volume v_p, n_- is the plasma density, and ν_i is the ionization frequency.

The power balance equation is approximately

$$fI_o V_o = P_i + I_- T_-, \tag{2}$$

where $fI_o V_o$ is the fraction of the beam power ($I_o V_o$) given to the plasma, P_i is the power lost in ionization and excitation processes, and $I_- T_-$ is the power lost because of electrons leaving the plasma volume, and T_- is the mean energy of the electrons in volts.

If we neglect excitation losses

$$P_i = \nu_i n_- e V_i v_p, \tag{3}$$

where V_i is the ionization potential. With the use of Eqs. 1 and 3, Eq. 2 becomes

$$I_+(T_- + V_i) = fI_o V_o. \tag{4}$$

In a typical experiment on system B with helium gas[2] used:

$f = 0.3$

$V_i = 24.6$ volts

$I_+ = 6.6$ amps

$I_o = 1$ amp

$V_o = 10,000$ volts.

Then

$$T_- \approx 430 \text{ volts}.$$

(b) The presence of high-energy plasma electrons has also been inferred from measurements of x-rays emitted by a molybdenum target. The target (1 in. × 1.5 in. × 0.01 in.) is movable, and experiments were performed with the target far from the beam (Fig. VII-1). At low pressures when a beam-plasma discharge cannot be sustained, no x-rays are observed. In the presence of the discharge x-rays are observed with the target parallel to the beam axis but the intensity is weak. With the target at 45° with respect to the beam axis, in either direction, the x-rays observed are considerably

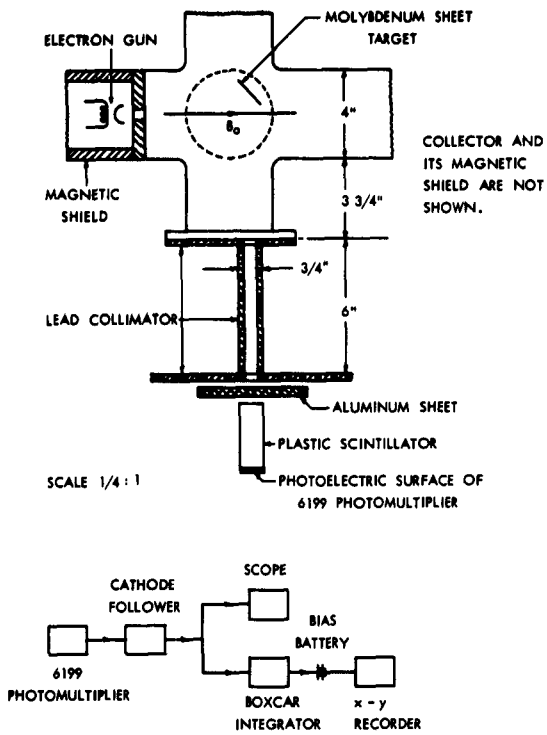


Fig. VII-1. Experimental geometry and schematic of x-ray absorption measurements in system A.

more intense. The results of a specific set of measurements are given in Fig. VII-2. These are for a beam voltage of 10 kv, with the target oriented 45° to the beam axis. By adjusting pressure and magnetic field for maximum x-ray intensity, we have been able to make absorption experiments with varying thicknesses of aluminum sheet up to 0.180 inch thick. The log of the observed intensities when plotted against the thickness of the absorber gave an average slope that falls between the K_β line of molybdenum (19.5 kv) and K_α line of tin (25.1 kv) as shown in Fig. VII-2. (The collimator resolution was such that both the target and the tin-coated wall were visible to the scintillator detector.) The measurement indicates the presence of electrons of approximately 20-kv energy or twice the beam voltage.

(c) The presence of high-energy electrons is also inferred from measurements of probe floating potential. The instantaneous voltage between a Langmuir probe and ground was measured at various positions in the discharge. The probe was grounded through a 1-megohm voltage divider. The time constant was approximately 1 μsec. A typical variation of floating potential with position at a given axial position is shown in

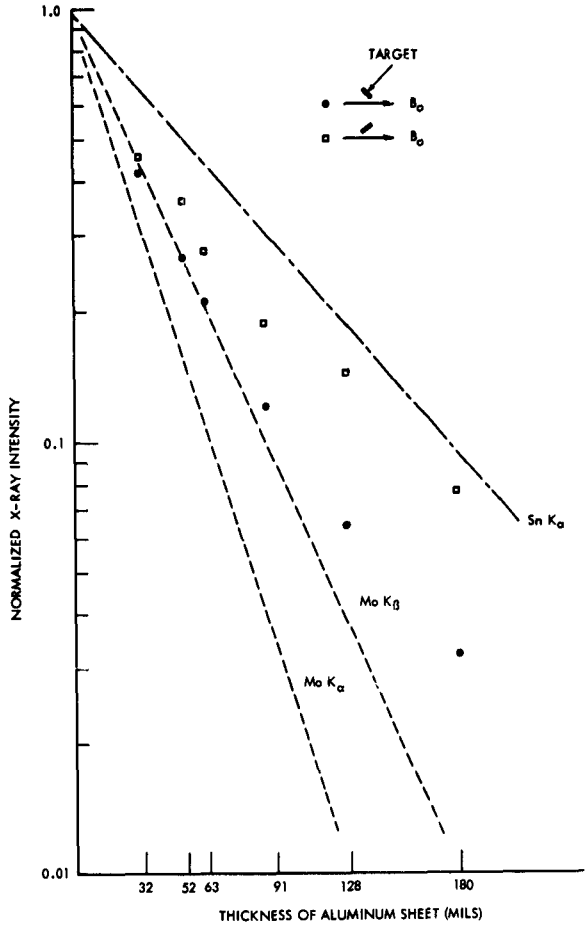


Fig. VII-2. Plot of normalized x-ray intensity vs thickness of aluminum sheet. Helium gas pressure, 0.2 μ Hg; magnetic mirror, 2:1; central field, 600 gauss; beam voltage, 10 kv.

Fig. VII-3. Negative floating potentials in excess of 1000 volts were observed when the beam voltage was 5.6 kv; this observation indicates the presence of high-energy electrons.

(d) Another estimate of the average electron energy was made by comparing the intensities of two He I lines. The light was observed just after the beam was turned off. Preliminary measurements indicate an electron temperature of 30-100 ev, but there are several conditions in the experiment which may not satisfy the assumed theoretical conditions.[3] Work will continue on this method to determine the validity of these preliminary results.

2. Electron Density

The plasma density ranges from 10^{11}/cc to 10^{13}/cc. This was estimated in two ways. First, the highest, strongest emitted frequency (that is not a harmonic of some lower emitted frequency) was assumed to equal ω_p, which is proportional to $n_-^{1/2}$. Second, the density was estimated from the average total charge flowing to the walls after the beam pulse had ended and from an estimate of the plasma volume.[2]

3. Plasma Ion Energy

In a typical helium beam-plasma discharge, the width of the 4686 Å line was found to be 0.4 Å. If this observed broadening were caused entirely by thermal motion of the ions, the ion temperature would be 5 ev. The plasma ion can acquire energy from the microwave fields, from the dc fields, or through collisional heating. The microwave and dc fields still have to be measured, but the fact that a positive floating potential is

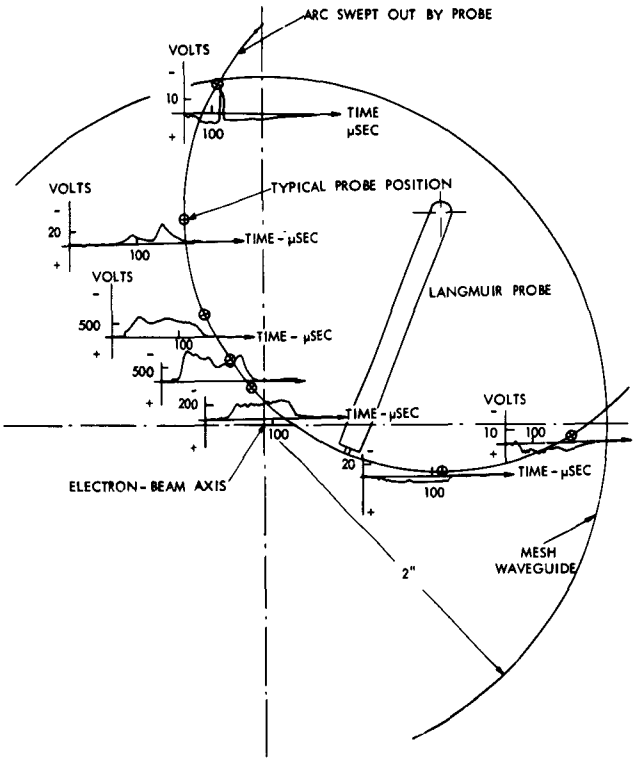


Fig. VII-3. Floating potential vs times at various positions in system A. Hydrogen gas pressure, 0.2 μ Hg; magnetic mirror, 2:1; central field, 600 gauss; beam voltage, 5.6 kv.

frequently observed near the walls (see Fig. VII-3) suggests that ions are expelled from the plasma by a large space potential between the discharge and the walls. This space-charge electric field may have been responsible for the spread of ion velocities inferred from the spectral broadening.

It can be shown that collisional heating of the ions by the electrons, competing with losses that are due to charge exchange and to loss of ions from the mirror, cannot produce an ion temperature of this magnitude, given our neutral particle density of 7×10^{12}/cc.

L. D. Smullin, W. D. Getty, B. A. Hartenbaum, H. Y. Hsieh

References

1. L. D. Smullin and W. D. Getty, Generation of a hot, dense plasma by a collective beam-plasma interaction, Phys. Rev. Letters 9, 3-6 (July 1, 1962).

2. L. D. Smullin, B. A. Hartenbaum, and H. Y. Hsieh, Electron beam plasma interaction experiments, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 71-76.

3. S. P. Cunningham, Spectroscopic Observations, U.S. AEC Technical Report WASH-289, University of California Radiation Laboratory, Livermore, California, June 1955, p. 279.

B. INSTABILITY IN THE HOLLOW-CATHODE DISCHARGE

An instability in the hollow-cathode discharge[1] has been observed which may be associated with the diffusion of plasma across the magnetic field away from the arc. We observe a single broad "spoke" of plasma rotating about the axis of the discharge. The density and shape of this cloud has been studied, and its behavior with various arc parameters is being observed. A theoretical study of the problem is being made.

1. Behavior of the Plasma Cloud

Our experimental arrangement is shown in Fig. VII-4. The arc is run on the magnetic axis of the system in argon at a pressure of several times 10^{-4} mm Hg. The magnetic field may be varied continuously up to approximately 600 gauss, and arc currents in the range 5-20 amps are used.

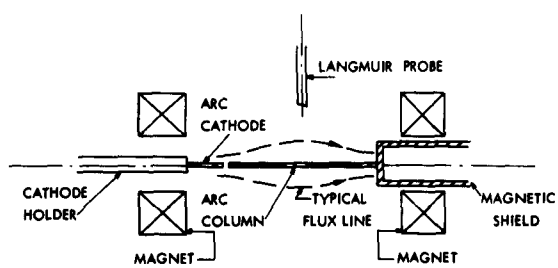


Fig. VII-4. Experimental arrangement.

The plasma cloud has been studied with a pair of movable Langmuir probes. The density and shape of the cloud has been approximately determined from ion saturation current and temperature, and the frequency of rotation of the cloud has been plotted against various parameters. We observe that the instability only appears for magnetic

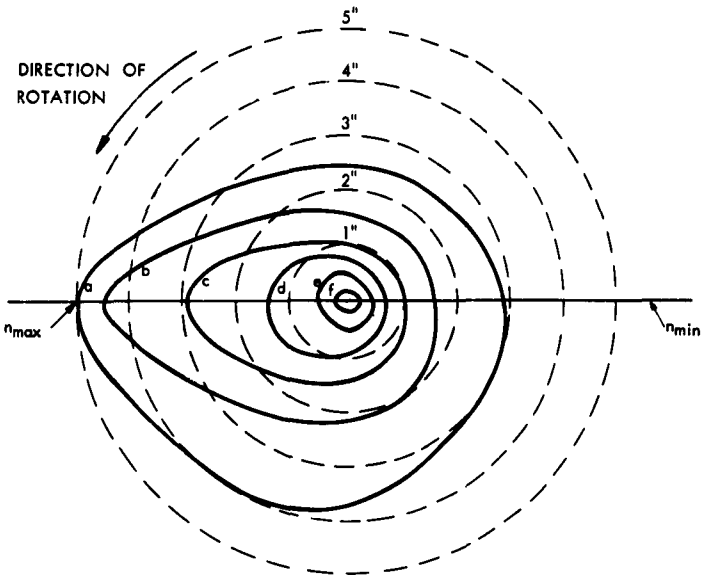


Fig. VII-5. Density Contour Map. Density in particles per cm^3: a, 10^{10}; b, 2×10^{10}; c, 4×10^{10}; d, 8×10^{10}; e, 1.8×10^{11}; f, 2.5×10^{11}. Magnetic field into the page.

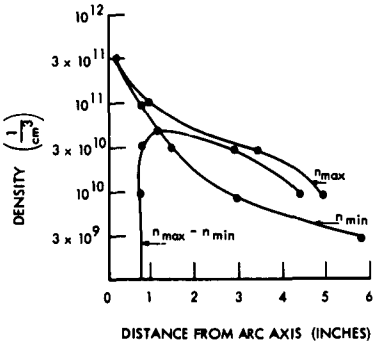


Fig. VII-6. Density profiles of disturbance.

fields above a certain value, which is not a strong function of pressure or arc current.

Figure VII-5 is a density contour map, each contour corresponding to a constant density at an instant of time. The whole map rotates counterclockwise at a few kilocycles. It is observed that the cloud does not exhibit any helical pattern, that is, there is no measurable phase shift of the density maximum with distance along the axis.

Figure VII-6 shows density profiles taken along the lines marked n_{max} and n_{min} on the contour map. Also, $n_{max}-n_{min}$ is plotted on the same axis.

Figure VII-7 illustrates the variations in rotational frequency with magnetic field and with arc current. The frequency is approximately proportional to the square root of magnetic field, and increases with increasing arc current. Figure VII-8 shows that the relation between the period of rotation and the background pressure is nearly linear.

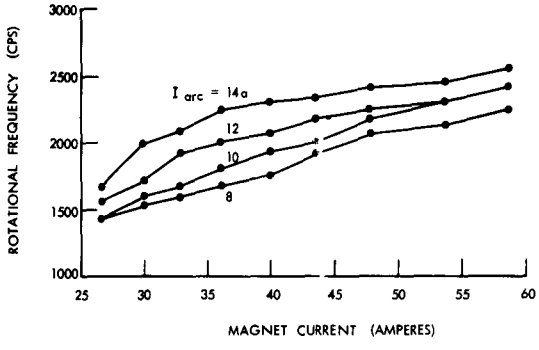


Fig. VII-7. Rotational frequency of disturbances vs magnet current, with arc current a parameter. Magnetic field on axis, approximately 10 gauss per ampere magnet current.

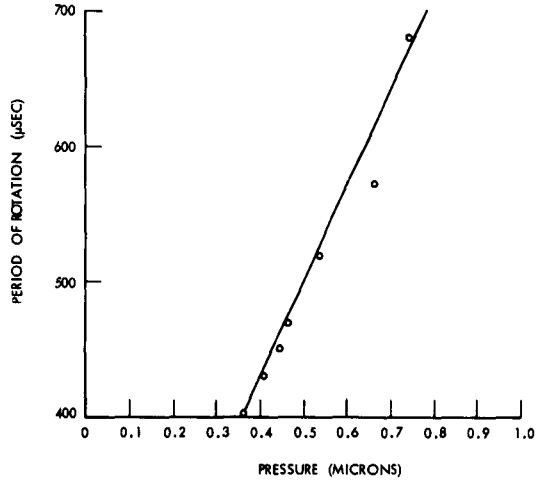


Fig. VII-8. Effect of background pressure on rotational period. Magnet current, 40 amps; arc current, 10 amps.

2. Theoretical Considerations

We postulate that the frequency of rotation of the cloud is such that the azimuthal velocity of the particles is equal to the (E_r/B_z) drift velocity. The direction of rotation agrees with this hyphothesis, since the plasma potential is positive with respect to the vacuum-chamber walls, thereby indicating an outward radial electric field. In a reflex discharge the plasma is contained between the cathode and a reflector, both between 50 and 80 volts negative with respect to the vacuum-chamber walls. In this case the sense of rotation of the cloud is observed to be opposite to that in the straignt (nonreflex) arc. The frequency of rotation is almost right for the values of E_r and B_z in the system, and has no relation to a cyclotron frequency of any type of particle present.

Although we have not yet established the mechanisms for this instability, we are trying to compare our observations with those predicted for the "flute" instability and for the helical instability of Kadomtsev and Nedospasov.[2]

D. L. Morse

References

1. D. L. Morse, Electron beam-plasma interaction experiments, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M.I.T., January 15, 1962, pp. 93-94.

2. B. B. Kadomtsev and A. V. Nedospasov, Instability of the positive column in a magnetic field and the 'anomalous' diffusion effect, J. Nuclear Energy, Part C, Vol. 1, pp. 230-235, 1960.

C. COUPLING OF EMPTY-WAVEGUIDE MODES, ELECTROSTATIC AND MAGNETOSTATIC MODES IN WAVEGUIDES LOADED WITH GYROTROPIC MEDIA

In this report we present a new approach to the solution of wave propagation in gyrotropic waveguides. Chorney used an argument based on coupling between an empty-waveguide mode and a quasi-static mode for the interpretation of a particular dispersion curve obtained by computations based on the exact determinantal equation.[1] Bers suggested the examination of the coupling of empty-waveguide and electrostatic modes for plasma-loaded waveguides. We now present the details of this approach and illustrate its use. Recently a similar coupling-of-modes approach was used in the solution of ferrite and plasma waveguides, as suggested by Auld.[2,3] However, the details of that approach are quite different from ours, as will be pointed out here.

1. Uncoupled Modes

Consider a gyrotropic medium characterized by dielectric and magnetic permittivity tensors

$$\overline{\overline{K}} = \begin{pmatrix} K_\perp & jK_x & 0 \\ -jK_x & K_\perp & 0 \\ 0 & 0 & K_\parallel \end{pmatrix} \tag{1}$$

$$\overline{\overline{L}} = \begin{pmatrix} L_\perp & jL_x & 0 \\ -jL_x & L_\perp & 0 \\ 0 & 0 & L_\parallel \end{pmatrix}. \tag{2}$$

We analyse $\overline{E}$ and $\overline{H}$ in a rotational and an irrotational part:

$$\overline{E} = \overline{E}^r - \nabla\Phi \tag{3a}$$

$$\overline{H} = \overline{H}^r - \nabla\Psi \tag{3b}$$

We expand $\overline{E}^r$ and $\overline{H}^r$ into series of empty-waveguide modes, and Φ and Ψ into series of electrostatic and magnetostatic modes. We then substitute these expansions in Maxwell's equations for the gyrotropic medium. The coupling of these modes because of the gyrotropic nature of the medium will then be established with the aid of the orthogonality relation for the individual modes. The orthogonality relations for electrostatic modes are

$$\gamma_\ell^2 \int K_\parallel \Phi_\ell \Phi_m^* \, da = \delta_{\ell m} \tag{4a}$$

$$\int \nabla_T \Phi_m^* \cdot \overline{\overline{K}}_T \cdot \nabla_T \Phi_\ell \, da = \delta_{\ell m}, \tag{4b}$$

and for the magnetostatic modes are

$$\gamma_m^2 \int L_\parallel \Psi_m \Psi_\ell^* \, da = \delta_{m\ell} \tag{5a}$$

$$\int \nabla_T \Psi_\ell^* \cdot \overline{\overline{L}}_T \cdot \nabla_T \Psi_m \, da = \delta_{m\ell}. \tag{5b}$$

The main difference with Auld and Eidson's treatment lies in the characterization of each quasi-static mode, not by its propagation constant (γ_m) but by other variables (plasma frequency in the case of plasma, radian frequency in the case of ferrite).

2. Coupled-Mode Theory

We define

$$\hat{E}_T^r = \sum_i V_i \hat{e}_{Ti} \tag{6a}$$

$$\hat{H}_T^r = \sum_i I_i \hat{h}_{Ti} \tag{6b}$$

$$\Phi = \sum_i F_i \Phi_i \tag{7}$$

$$\Psi = \sum_i M_i \Psi_i. \tag{8}$$

Substituting Eqs. 6-8 in Maxwell's equations and applying the orthogonality relations, we obtain

$$\gamma(\bar{V}+\bar{\bar{a}}\,\bar{F}) = \bar{\bar{b}}\,\bar{I} + \bar{\bar{c}}\,\bar{M} \tag{9a}$$

$$\gamma(\bar{I}+\bar{\bar{d}}\,\bar{M}) = \bar{\bar{e}}\,\bar{V} + \bar{\bar{f}}\,\bar{F} \tag{9b}$$

$$\bar{F} = \frac{\bar{\bar{f}}^\dagger}{j\omega\epsilon_o}\bar{V} + \frac{\gamma\bar{\bar{a}}^\dagger}{j\omega\epsilon_o}\bar{I} \tag{10}$$

$$\bar{M} = \frac{\bar{\bar{c}}^\dagger}{j\omega\mu_o}\bar{I} + \frac{\gamma\bar{\bar{d}}^\dagger}{j\omega\mu_o}\bar{V}. \tag{11}$$

Here, the dagger means complex conjugate of the transpose of a matrix, and

$$a_{i\ell} = p_{ei}^2 \int \phi_i^* \Phi_\ell \, da \tag{12a}$$

$$b_{in} = \frac{1}{j\omega\epsilon_o} p_{en}^2 p_{ei}^2 \int \frac{\phi_i^* \phi_n}{K_\parallel} da + j\omega\mu_o \int \hat{h}_{Ti}^* \cdot \bar{\bar{L}}_T \cdot \hat{h}_{Tn} \, da \tag{12b}$$

$$c_{im} = j\omega\mu_o \int \hat{h}_{Ti}^* \cdot \bar{\bar{L}}_T \cdot (-\nabla_T \Psi_m) \, da \tag{12c}$$

$$d_{im} = p_{hi}^2 \int \psi_i^* \cdot \Psi_m \, da \tag{12d}$$

$$e_{in} = \frac{1}{j\omega\mu_o} p_{hi}^2 p_{hn}^2 \int \frac{\psi_i^* \psi_n}{L_\parallel} da + j\omega\epsilon_o \int \hat{e}_{Ti}^* \cdot \bar{\bar{K}}_T \cdot \hat{e}_{Tn} \, da \tag{12e}$$

$$f_{i\ell} = j\omega\epsilon_o \int \hat{e}_{Ti}^* \cdot \bar{\bar{K}}_T \cdot (-\nabla_T \Phi_\ell) \, da. \tag{12f}$$

Lower-case letters for potentials denote the empty-waveguide modes; and upper-case letters, the quasi-static modes. The symbol p denotes transverse wave numbers.

3. Discussion

If we consider all of the empty-waveguide modes, the quasi-static modes (Eqs. 10 and 11) are eliminated, and we get the equations of Chorney.[1] This is to be expected,

since the empty-waveguide modes form a complete set. When we do not consider the complete set of empty-waveguide modes, the first two of Maxwell's equations are satisfied only approximately, and the divergence equations must be satisfied independently; hence Eqs. 10 and 11 are then independent of Eqs. 9a and 9b.

In the case of a waveguide completely filled by a dielectric medium each electrostatic potential is related to only one transverse electromagnetic potential $\left(\Phi_\ell = \frac{1}{\sqrt{K_\perp}}\phi_\ell\right)$. Therefore the quasi-static modes are eliminated in our treatment. This does not happen for ferrite-loaded waveguides. There is coupling of each magnetostatic potential to more than one transverse electromagnetic potential, which is due to the boundary conditions.

Our treatment is, however, most useful for partially filled waveguides. We assume that the characteristic tensors are different from 1 in the space occupied by the gyrotropic medium, and equal to 1 in the free space outside the medium. The formulation given here allows for the tensors to be functions of the transverse coordinates, and hence the results also apply directly to the partially filled waveguide.

4. Example

For a completely filled circular cylindrical plasma waveguide of radius a we examine the coupling among the TE_{10} mode, TM_{10} mode, and the $Q.S._1$ mode. There exists a magnetostatic field along the axis of the waveguide. The propagation constant is determined by the determinant of Eqs. 9 and 10. The dominant term in the expansion of this determinant contains the three assumed modes and leads to the following approximate expression for the propagation constant:

$$\gamma^2 = -u^2 K_\perp + \frac{1}{2}\left(p_{h1}^2 + \frac{K_\perp}{K_\parallel} p_{e1}^2\right) \pm \frac{1}{2}\left[\left(\frac{K_\perp}{K_\parallel} p_{e1}^2 - p_{h1}^2\right)^2 + 3.2\,\frac{K_x^2}{K_\parallel} u^2\left(K_\parallel u^2 - p_{e1}^2\right)\right]^{1/2} \tag{13}$$

The frequencies are normalized over $\omega_N = c/a$ $(u = \omega/\omega_N)$, and the propagation constants over $k_N = 1/a$.

Equation 11 gives the following expressions for cutoffs $(u_{ec}; u_{hc})$:

$$u_{ec}^2 - u_p^2 = p_{e1}^2 \qquad \text{(E-cutoffs)} \tag{14a}$$

$$u_{hc}^2 \frac{K_\perp^2 - 0.8K_x^2}{K_\perp} = p_{h1}^2 \qquad \text{(H-cutoffs).} \tag{14b}$$

The only difference from the exact expressions is the numerical coefficient 0.8 that appears in Eq. 14b.

For the resonances we find

Cyclotron resonances ($u = u_{ci}$ or u_{ce})

$$\frac{\gamma^2}{K_\perp} \longrightarrow -1.9u^2 + \frac{p_{el}^2}{K_\parallel}. \tag{15}$$

In the exact solution[4] we have 2.0 instead of 1.9.

Plasma resonance ($u = u_{p0}$)

$$\gamma^2 K_\parallel \longrightarrow p_{el}^2 K_\perp \tag{16}$$

which is identical with the exact solution.[4]

These expressions include the change from forward waves to backward waves at cyclotron resonance.

P. E. Serafim, A. Bers

References

1. P. Chorney, Sc.D. Thesis, Department of Electrical Engineering, M.I.T., September 1961.

2. B. A. Auld, Internal Memorandum M. L. No. 931, Microwave Laboratory, Stanford University, June 1962.

3. B. A. Auld and J. C. Eidson, Internal Memorandum M. L. No. 964, Microwave Laboratory, Stanford University, October 1962.

4. W. P. Allis, S. J. Buchsbaum, and A. Bers, Waves in Anisotropic Plasmas (The M.I.T. Press, Cambridge, Mass., 1963).

D. ION OSCILLATIONS IN NEUTRALIZED, THIN ELECTRON BEAMS

In this report we present a possible theoretical explanation of the mechanism for inducing ion oscillations in thin electron beams. The instability is of a "hose" type ($n = \pm 1$) in which the ac velocity of the beam is largely in the transverse direction.

Consider a thin cylindrical electron beam of radius b partially or fully neutralized by an ion background of the same radius. It is assumed that there are no stationary electrons. The beam moves with dc velocity v_o parallel to a magnetic field B_o, both in the z direction. We assume that the metal drift-tube radius is much larger than b.

For such a thin beam, a quasi-static treatment is justified when the beam is non-relativistic. We consider the angular dependences $e^{jn\phi}$ with $n = \pm 1$, and look for wave solutions of the form $e^{j(\omega t - \beta z)}$. If we take the small-argument limit of the appropriate

Bessel functions, the potentials in the vicinity of the beam are

$$\phi_i = K\frac{r}{b}e^{\pm j\phi} \qquad \text{inside}$$
$$\phi_o = K\frac{b}{r}e^{\pm j\phi} \qquad \text{outside.} \tag{1}$$

If we equate the discontinuity in the normal electric field to the surface charge, we obtain the following dispersion equation

$$\frac{\omega_{pb}^2}{\omega_r(\omega_r \pm \omega_{ce})} + \frac{\omega_{pi}^2}{\omega(\omega \mp \omega_{ci})} = 2, \tag{2}$$

where $\omega_r = \omega - \beta v_o$, ω_{pb} is the beam-plasma frequency, ω_{pi} is the ion-plasma frequency, ω_{ce} is the electron-cyclotron frequency, and ω_{ci} is the ion-cyclotron frequency. The beam waves in the absence of the ions are shown in Fig. VII-9. We see that for

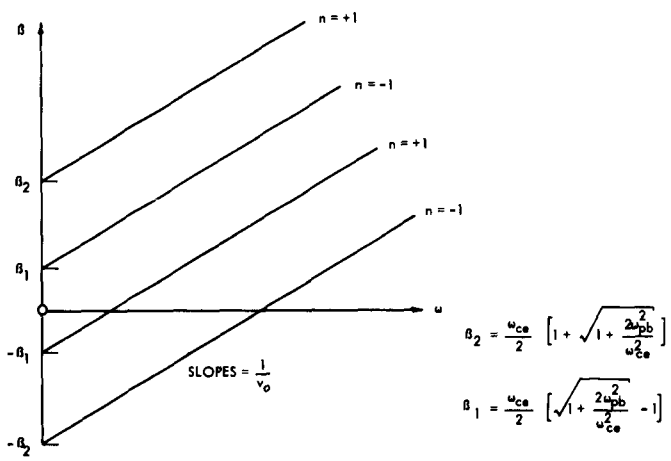


Fig. VII-9. Electron-beam waves in the absence of ions.

$\omega_{pb} \ll \omega_{ce}$, we have the usual cyclotron and synchronous waves of a filamentary beam. The dispersion in the presence of the ions is sketched in Fig. VII-10 for $\omega_{pi} \gg \omega_{ci}$. We see that there are convective instabilities for both $n = \pm 1$ which have an infinite rate of growth in space for $\omega = \omega_{pi}/\sqrt{2}$. This instability can be interpreted as a coupling of the "generalized" cyclotron and synchronous waves because of the ions. It is a type of reactive-medium amplification in the sense that the ion "oscillators" appear inductive just below the resonant frequency $\omega_{pi}/\sqrt{2}$.

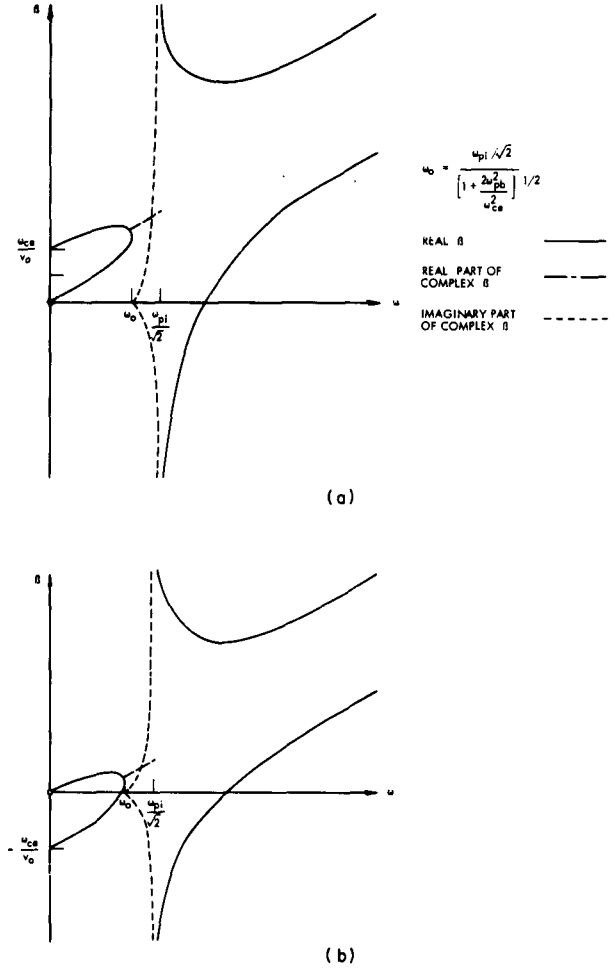


Fig. VII-10. Electron beam-ion interaction.
(a) n = +1; (b) n = −1.

We should note that the infinite gain violates the assumption of $\beta b \ll 1$; however, a gain that is comparable to 1/b is extremely large.

R. J. Briggs

E. ELECTRON-CYCLOTRON PLASMA HEATING

Investigations on the production of hot dense plasmas by microwaves at the electron-cyclotron frequency (Quarterly Progress Report No. 64, pages 103-104, and No. 65,

pages 95-97) have been resumed. An experiment at medium power is being carried out (Fig. VII-11); it utilizes an RK 62 pulsed magnetron at λ = 10 cm, and a cylindrical

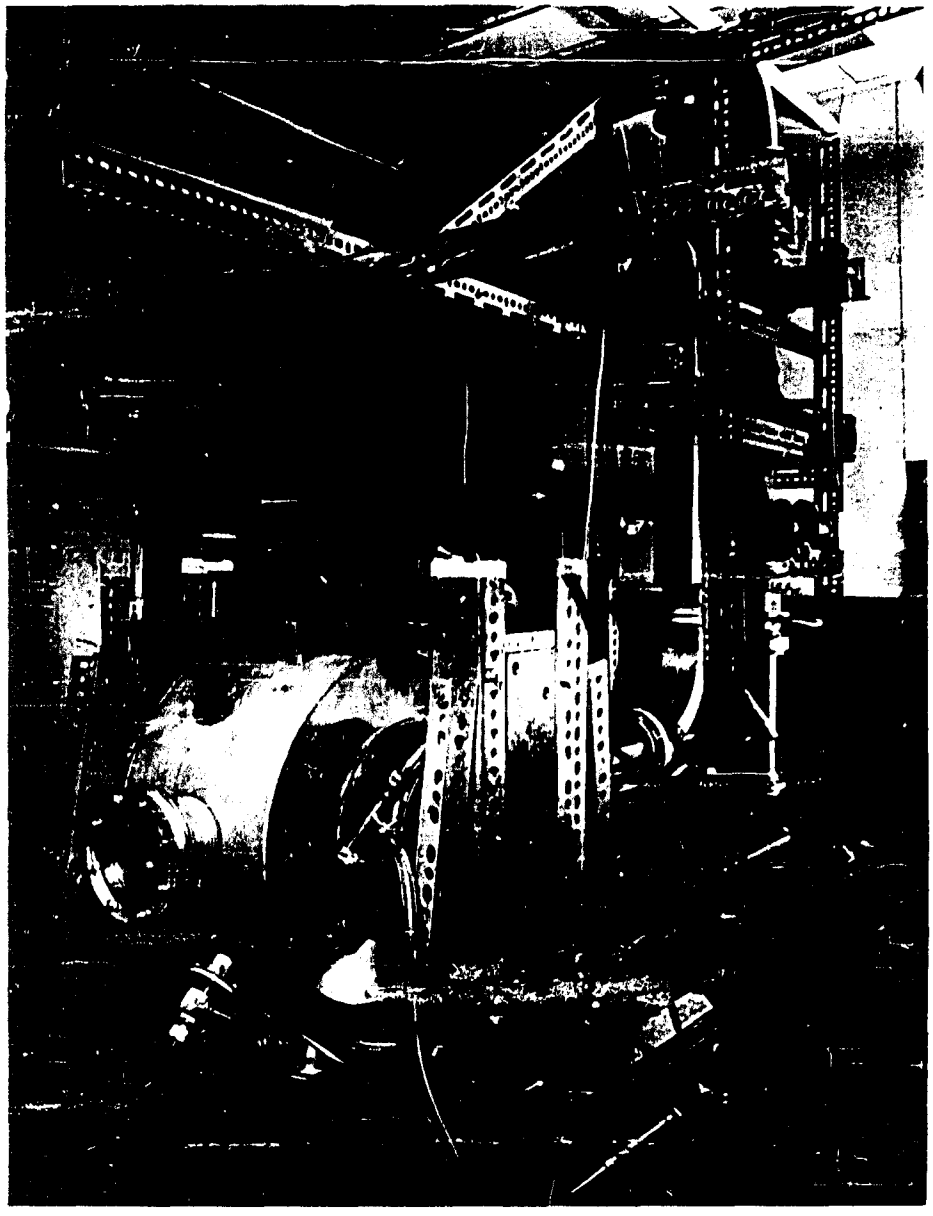


Fig. VII-11. Cavity and waveguide assembly.

cavity with concave end plates. This cavity is made of copper-plated stainless steel and has a volume of approximately 25 liters, so that it has a high mode density in the neighborhood of 3000 mc. Preliminary experiments were made with 1-μsec pulses at

a 500/sec repetition rate, and available peak power of 50-60 kw produced visible discharges at air pressures of approximately 10^{-4} mm Hg, with x-ray intensity of tens of kilovolts.

G. Fiocco, L. D. Smullin

F. ELECTRON-BEAM INTERACTIONS WITH IONS IN A WARM PLASMA

A theoretical study is being made of the various beam-plasma instabilities in a warm plasma in order to determine the feasibility of heating ions directly from a collective type of interaction. In this report we summarize some results of a one-dimensional analysis of the problem.

We consider the one-dimensional system of an unbounded electron beam moving with velocity v_o through an unbounded plasma parallel to an applied steady magnetic field, B_o (both taken to be in the z direction). The waves along the magnetic field have the dependence $\exp[j(\omega t-kz)]$; these waves split into the longitudinal and transverse waves, which are uncoupled and therefore can be considered separately.

1. Longitudinal Interaction

The longitudinal waves have only a longitudinal component of the electric field and zero magnetic field. The dispersion equation can be written

$$\frac{\omega_{pb}^2}{(\omega-kv_o)^2} = K_{\parallel}(\omega, k), \tag{1}$$

where ω_{pb} is the beam-plasma frequency and $K_{\parallel}$ is the longitudinal dielectric constant of the plasma. From a collisionless kinetic treatment, assuming cold ions, one obtains[1]

$$K_{\parallel} = 1 - \frac{\omega_{pi}^2}{\omega^2} - \omega_{pe}^2 \int_{-\infty}^{\infty} \frac{f_o(v_z)\, dv_z}{(\omega-kv_z)^2}, \tag{2}$$

where f_o is the unperturbed velocity distribution and ω_{pi} and ω_{pe} are the ion- and electron-plasma frequencies, respectively. From a transport equation model, assuming an isotropic pressure, one has[1]

$$K_{\parallel} = 1 - \frac{\omega_{pi}^2}{\omega^2} - \frac{\omega_{pe}^2}{\omega^2 - k^2 V_{Te}^2}, \tag{3}$$

where V_{Te} is the average thermal velocity of the plasma electrons. It can be shown that the two formulations give identical results when (ω/k) is either much larger or much

smaller than V_{Te}, even though they are derived from quite different physical assumptions.

a. Weak Beam

For a beam density that is much less than the plasma density, to solve for the beam waves we can set $k = (\omega/v_o) + \delta$ and assume that $\delta \ll \omega/v_o$. Therefore,

$$K_{||}(\omega, k) \approx [K_{||}]_{\omega=kv_o} + \left[\frac{\partial K_{||}}{\partial k}\right]_{\omega=kv_o} \delta . \tag{4}$$

A necessary, but not sufficient, condition for the validity of this approximation is that $(n_b/n_p)(T_e/2V_o) \ll 1$, where n_b and n_p are the beam and plasma densities, T_e is the electron temperature, and V_o is the voltage corresponding to the beam velocity v_o. For $T_e/V_o \ll 1$, much stronger conditions on n_b/n_p are required.

For the transport equation model, Eq. 3, we see that a convective instability is obtained in the regions in which $[K_{||}]_{\omega=kv_o}$ is negative, and that the maximum spacial growth rate occurs near the synchronism of the plasma wave with the beam ($[K_{||}]_{\omega=kv_o} = 0$). The mechanism here is the well-known reactive-medium amplification: when a bunched electron beam passes through a medium with a negative dielectric constant (an "inductive medium"), the electrons in a bunch attract rather than repel and hence the bunching is further enhanced. In Fig. VII-12 the regions in which $K_{||} < 0$ are plotted in the ω-k plane. We see that for $2V_o > T_e$ the maximum amplification occurs above

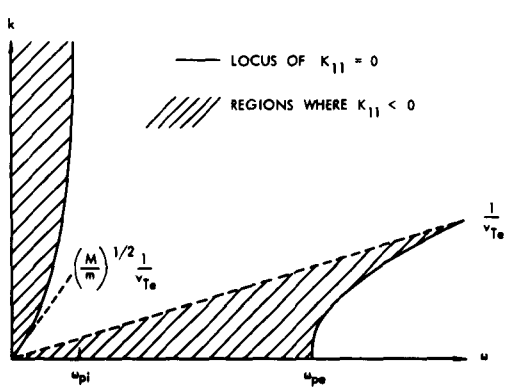


Fig. VII-12. Regions in which $K_{||} < 0$.

ω_{pe}. This amplification has been considered by a number of authors.[2] For $(m/M)T_e < 2V_o < T_e$, there is no instability. For $2V_o < (m/M)T_e$, there is again amplification, which now is a maximum at a frequency less than ω_{pi}; the growth rate in space is

$$\text{Im } k = \frac{3^{1/2}}{2^{4/3}} \frac{\omega_{pi}}{v_o} \left(\frac{n_b}{n_p} \frac{T_e}{2V_o} \right)^{1/3} \left(1 - \frac{M}{m} \frac{2V_o}{T_e} \right)^{1/2}, \tag{5}$$

which is less than ω_{pi}/v_o for the conditions assumed here.

The collisionless kinetic treatment, Eq. 2, gives similar results except in the region in which $2V_o$ is comparable to T_e. The dielectric constant $[K_{||}]_{\omega=kv_o}$ is now complex; it can be shown that this fact leads to a convective instability even in the region in which $2V_o < T_e$. The mechanism for the instability is now of a "resistive-medium" type[3]; the novel feature in the present analysis is that the loss in the medium comes about because of the Landau damping. This amplification is plotted in Fig. VII-13 for a Maxwellian

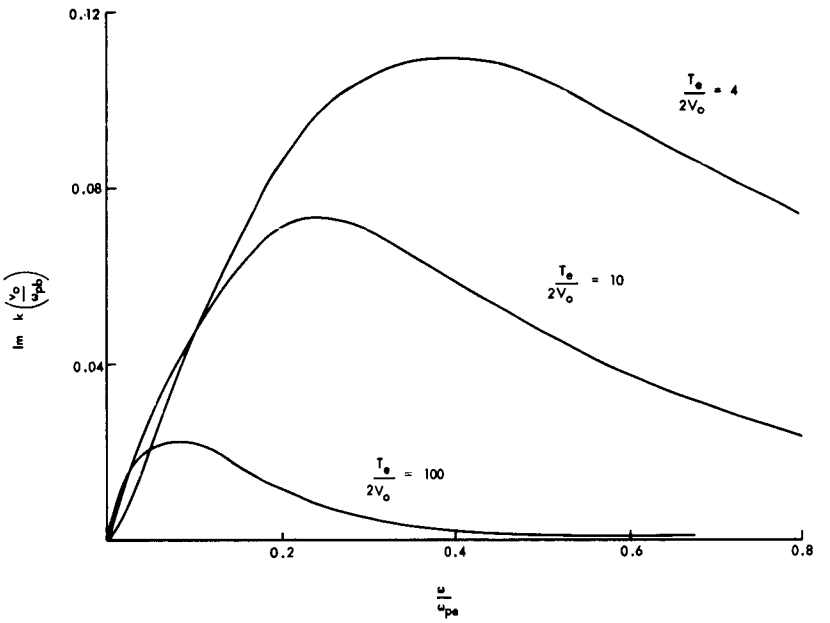


Fig. VII-13. Resistive-medium amplification.

distribution of velocities. The frequency of maximum gain is of the order of $(2V_o/T_e)^{1/2}$ ω_{pe} for $2V_o < T_e$; however, the peak of the gain decreases rapidly if $2V_o/T_e$ is decreased, since the Landau damping is then much less.

b. Strong Beam

If the beam density is large enough so that $(n_b/n_p)(T_e/2V_o) > 1$, a very interesting transition takes place. This transition can be anticipated by looking at $k \to \infty$ as $\omega \to \omega_{pi}$;

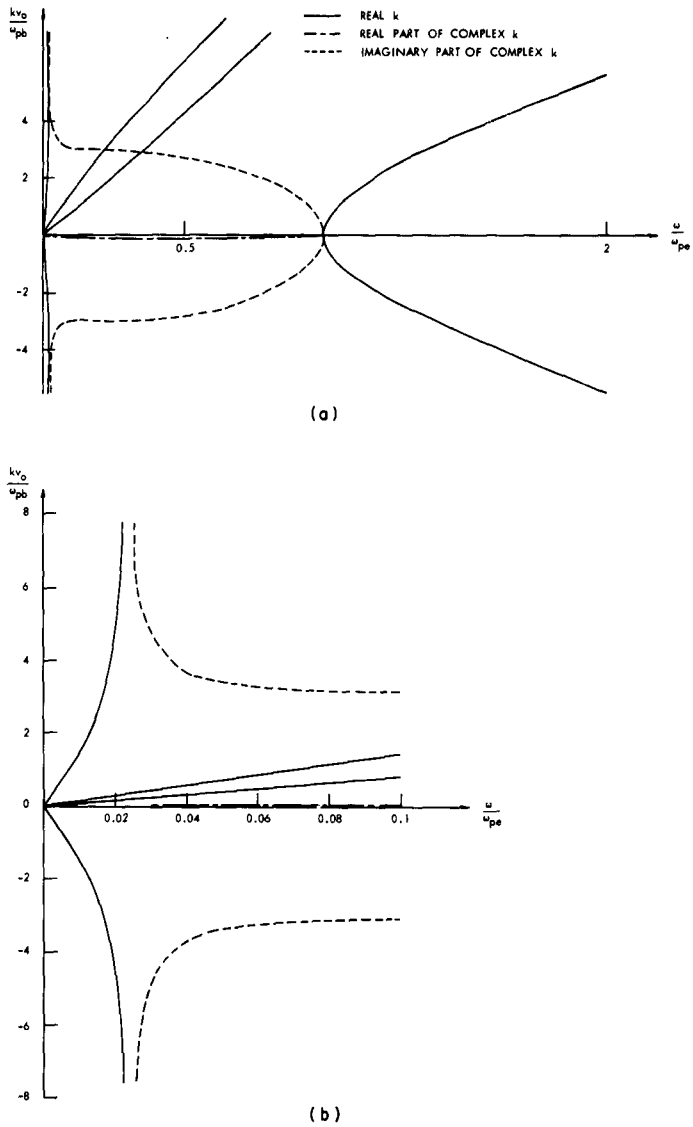
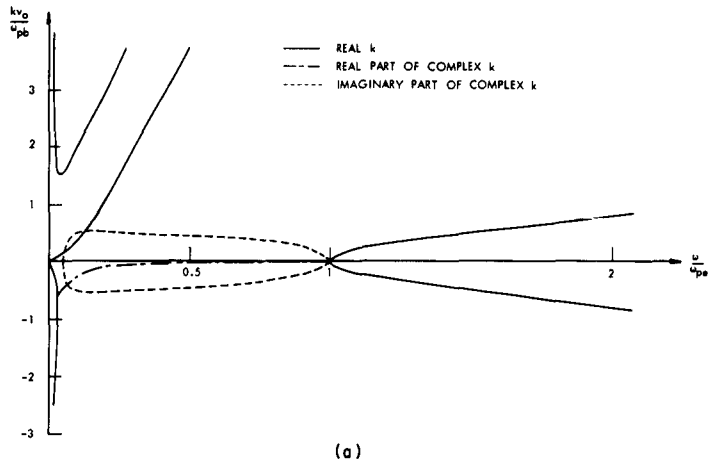
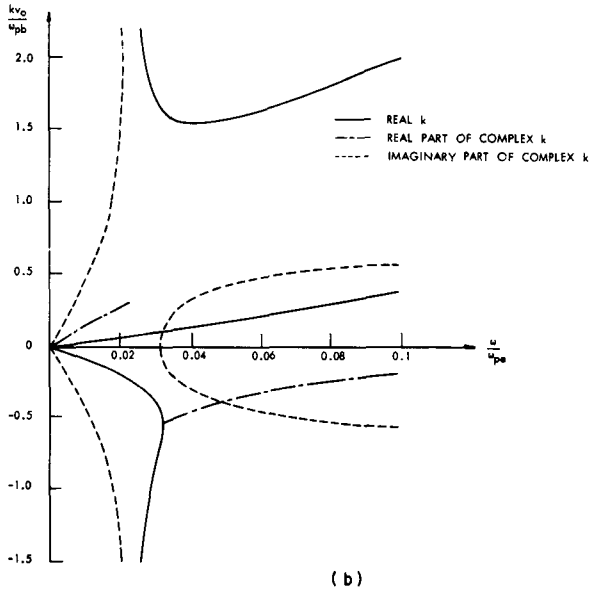


Fig. VII-14. (a) Longitudinal dispersion $n_b/n_p = 10^{-2}$, $T_e/2V_o = 10$.
(b) Detail at low frequency.



(a)



(b)

Fig. VII-15. (a) Longitudinal dispersion $n_b/n_p = 10^{-2}$, $T_e/2V_o = 400$.
(b) Detail at low frequency.

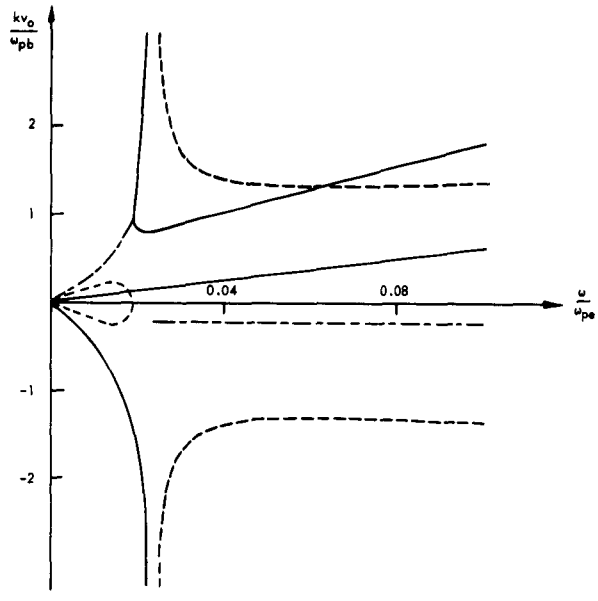


Fig. VII-16. Longitudinal dispersion $n_b/n_p = 10^{-2}$, $T_e/2V_o = 50$.

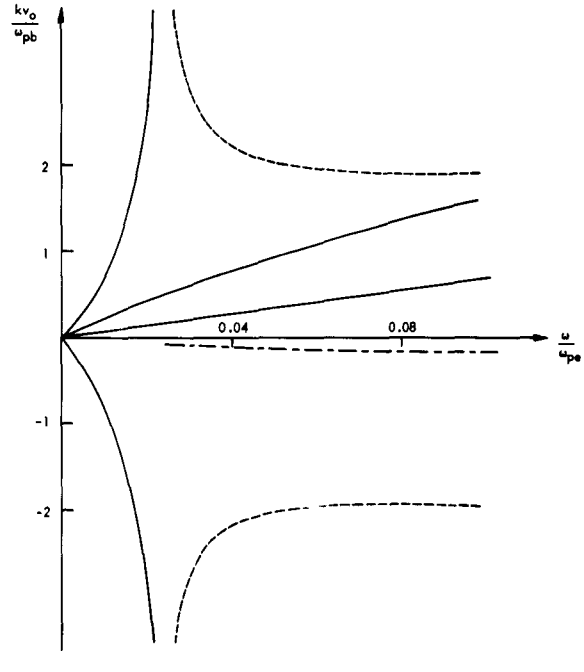


Fig. VII-17. Longitudinal dispersion $n_b/n_p = 10^{-2}$, $T_e/2V_o = 25$.

$$\frac{1}{k^2}\left[\frac{\omega_{pe}^2}{V_{Te}^2} - \frac{\omega_{pb}^2}{v_o^2}\right] = \frac{\omega_{pi}^2}{\omega^2} - 1. \tag{6}$$

We see that for $(n_b/n_p)(T_e/2V_o) > 1$, real k with $k \to \infty$ is obtained for $\omega > \omega_{pi}$ rather than $\omega < \omega_{pi}$. In Fig. VII-14, the complete dispersion is presented with the pressure formulation used for simplicity and with a proton plasma having the system parameters $n_b/n_p = 10^{-2}$ and $(n_b/n_p)(T_e/2V_o) = 0.1$. In Fig. VII-14, ω is real for all real k and the complex roots of k represent evanescent waves. In Fig. VII-15 the dispersion is plotted for $(n_b/n_p)(T_e/2V_o) = 4$; we see that now complex ω for real k is obtained (by counting the real roots), and the complex roots of k occurring for $\omega \leq \omega_{pi}$ now represent a convective instability with the spacial growth rate tending to infinity at $\omega = \omega_{pi}$.

If $(n_b/n_p)(T_e/2V_o)$ is only slightly less than one, there is still a convective instability but the spacial rate of growth is no longer infinite. This transition is illustrated in Figs. VII-16 and VII-17, in which the low-frequency dispersion is plotted for $(n_b/n_p)(T_e/2V_o) = 0.5$ and 0.25. We see that the instability has disappeared in the latter case (Fig. VII-17).

The condition for infinite gain at ω_{pi} can also be written as $\lambda_{pb} < \lambda_D$, where $\lambda_{pb} = 2\pi v_o/\omega_{pb}$ is the space-charge wavelength and $\lambda_D = 2\pi V_{Te}/\omega_{pe}$ is the Debye length in the plasma. This form is physically appealing, since it suggests that the "beam wavelength" should fit inside a Debye sphere in order to have strong interaction with the ions.

2. Transverse Interactions

The dispersion equation for the transverse waves splits into separate equations for right and left circularly polarized waves. From a collisionless kinetic treatment, assuming an isotropic equilibrium velocity distribution, one obtains

$$\frac{k^2c^2}{\omega^2} = 1 - \sum \frac{\omega_p^2}{\omega}\int \frac{f_o(\bar{v})\, d^3v}{\omega - kv_z \mp \omega_c} - \frac{\omega_{pb}^2(\omega - kv_o)}{\omega^2(\omega - kv_o \pm \omega_{ce})}, \tag{7}$$

where the sum is over all species of the plasma, and ω_{ce} and ω_{ci} are the electron- and ion-cyclotron frequencies, respectively. For a relativistic beam, the relativistic (transverse) mass should be used in computing ω_{pb} and the ω_{ce} of the beam, that is,

$$m = \frac{m_o}{\left(1 - \frac{v_o^2}{c^2}\right)^{1/2}}, \tag{8}$$

where m_o is the rest mass. It should also be noted that the inclusion of temperature

by means of the transport equations with an isotropic pressure would predict no effect of temperature, since there is no bunching associated with the transverse waves.

a. Cold Plasma

The transverse wave interactions in a cold plasma were considered in Quarterly Progress Report No. 67 (pages 80-83). In that report, the conditions for charge neutrality were not carefully considered; consideration of these conditions here leads to a slight change in the very low frequency nonconvective instability. For a cold plasma, Eq. 7 becomes

$$\frac{k^2c^2}{\omega^2} = 1 - \frac{\omega_{pe}^2}{\omega(\omega\pm\omega_{ce})} - \frac{\omega_{pi}^2}{\omega(\omega\mp\omega_{ci})} - \frac{\omega_{pb}^2(\omega-kv_o)}{\omega^2(\omega-kv_o\pm\omega_{ce})}. \tag{9}$$

Requiring that the ion density in the plasma just equal the beam density plus the plasma electron density, and assuming that $\omega \ll \omega_{ci}$ and $kv_o \ll \omega_{ce}$, for the left polarized wave we obtain for Eq. 9

$$\omega^2 = k(k-k_o)\, u_a^2, \tag{10}$$

where u_a is the Alfvén speed in the plasma, $k_o = \mu_o J_{ob}/B_o$, J_{ob} is the beam current density, and we have assumed that $u_a \ll c$. This expression is a valid approximation for $k \leq k_o$ if $k_o \ll \omega_{ce}/v_o$, which requires that $\left(\omega_{pb}^2/\omega_{ce}^2\right)\left(v_o^2/c^2\right) \ll 1$. This dispersion is plotted in Fig. VII-18; we see that complex ω is obtained for real k between zero and k_o, whereas k is real for all real ω in this region; these results indicate a nonconvective instability. It can be shown that for such an instability to be excited, the length of the system much exceed a certain critical "starting length." For perfect electric conductors terminating the system, the shortest starting length is given by

$$L_{start} = \frac{2\pi}{k_o} = \frac{2\pi B_o}{\mu_o J_o}. \tag{11}$$

It is interesting to note also that this instability corresponds to a pure exponential growth in time of an initial spacial pattern of particle velocities.

An analysis of the right polarized wave shows that a similar nonconvective instability is obtained, only with $-k_o$ appearing in Eq. 10. It is easily shown that this represents exactly the same physical phenomena.

b. Warm Plasma

It can be shown that the electron temperature has essentially no effect on the convective instability near ω_{ci}, as would be expected physically. From Eq. 7 we see that for the ion temperature to be neglected it is necessary that $kV_{Ti} \ll \omega_{ci} - \omega$ near the

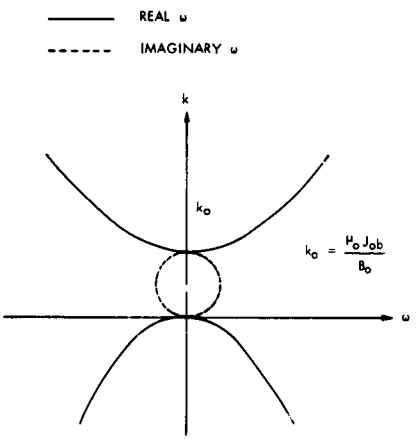


Fig. VII-18. Low-frequency transverse instability (left polarized).

intersection of the plasma wave and the beam cyclotron wave, where V_{Ti} is the average thermal velocity of the ions. This requires that

$$\frac{MV_{Ti}^2}{mv_o^2} \ll \frac{m}{M}\left(\frac{v_o}{c}\right)^4\left(\frac{\omega_{pi}}{\omega_{ce}}\right)^4. \tag{12}$$

(As an example, for protons of density 10^{12}/cc, magnetic field of 1000 gauss, and v_o/c = 0.2, Eq. 12 requires that the ion temperature be less than 10^{-7} volts, or much less than 1°K.) We conclude that for most cases of practical interest the ion temperature cannot be neglected in discussing the interaction at the ion-cyclotron frequency.

One finds, however, that there is still a convective instability near $\omega = \omega_{ci}$ when the ion temperature is included in Eq. 7. This interaction is again of a "resistive-medium" type, the physical origin of the "loss" mechanism now being the cyclotron damping in the plasma. For reasonable ion temperatures, such that the inequality of Eq. 14 is reversed, the maximum growth rate in space is

$$\text{Im } k = \sqrt{\frac{\pi}{2}}\frac{\omega_{ci}}{V_{Ti}}\frac{\omega_{pi}^2\omega_{pb}^2}{\omega_{ce}^4}\left(\frac{v_o}{c}\right)^4. \tag{13}$$

(For the example cited below Eq. 14, by assuming a beam density of 10^{10}/cc and ions of 0.1 ev, this growth rate is of the order of 10^{-5}/cm, which is a very weak amplification.)

The effect of temperature on the nonconvective instability, Eq. 10, is quite small, as would be expected, since the frequency is far from resonance. To neglect temperature

effects, it is sufficient to require (see Eq. 7) that kV_{Te}/ω_{ce} and kV_{Ti}/ω_{ci} be much less than one for the maximum wave number that must be considered. This requires that

$$\frac{T_i}{2V_o} = \frac{MV_{Ti}^2}{mv_o^2} \ll \frac{M}{m}\left(\frac{u_a}{v_o}\right)^4\left(\frac{n_p}{n_b}\right)^2 \tag{14}$$

$$\frac{T_e}{2V_o} = \frac{mV_{Te}^2}{mv_o^2} \ll \left(\frac{M}{m}\right)^2\left(\frac{u_a}{v_o}\right)^4\left(\frac{n_p}{n_b}\right)^2. \tag{15}$$

(As an example, for protons in a magnetic field of 1000 gauss, beam density of 10^{10}/cc, and $V_o = 10^4$ volts, inequalities (14) and (15) are satisfied for ion temperatures less than 4×10^5 volts and electron temperatures less than 8×10^8 volts.)

It might be expected that collisions would have an important effect on this nonconvective instability, since ω is so small; however, it can be shown that the starting length is not altered if a relaxation type of collision frequency is included, although the peak growth rate in time is reduced.

R. J. Briggs, A. Bers

References

1. See, for example, the following review articles: F. W. Crawford and G. S. Kino, Proc. IRE 49, 1767-1789 (1961); Ya. B. Fainberg, J. Nuclear Energy, Part C 4, 203 (1962).

2. See, for example, A. I. Akhiezer and Ya. B. Fainberg, Soviet Phys. – JETP 21, 1262-1269 (1951); G. D. Boyd, L. M. Field, and R. W. Gould, Proc. Symposium on Electronic Waveguides (Polytechnic Institute of Brooklyn, Brooklyn, New York, 1958), pp. 367-375; V. S. Imshennik and Yu. I. Morozov, Soviet Phys. – Tech. Phys. 6, 464-471 (1961); and M. A. Lampert, J. Appl. Phys. 27, 5-11 (1956).

3. C. K. Birdsall, G. R. Brewer, and A. V. Haeff, Proc. IRE 41, 865-875 (1953).

G. SCATTERING OF LIGHT FROM (PLASMA) ELECTRONS III

By using slight improvements in the techniques developed in a previously reported electron-beam experiment,[1,2] we have observed scattered optical radiation from the electrons in a dc thermal plasma. Development of these techniques will provide a powerful plasma diagnostic tool, since it is possible to obtain detailed measurements of the velocity distribution and density of the plasma electrons. These measurements can be made with very high spatial resolution (~1 mm^3); also, time resolution of these quantities should not present any great problems. Furthermore, this information is obtained without the introduction of probes, or, as is the case in some spectroscopic

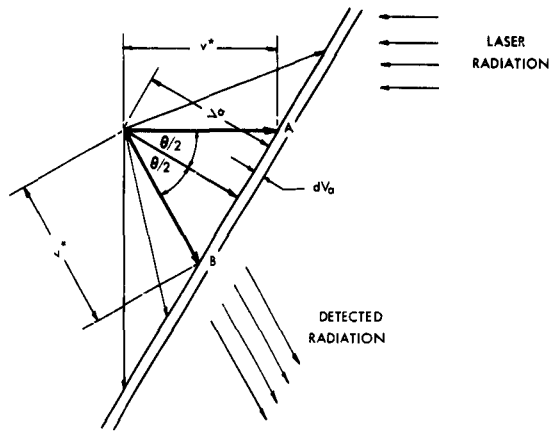


Fig. VII-19. Diagram used in derivation of spectrum of scattered radiation.

measurements, any impurity gases.

The derivation of the expected spectrum is made as follows. Consider an electron in velocity space at A with velocity equal to v^* (see Fig. VII-19); then this electron will "see" an input wavelength from the laser which is Doppler-shifted by an amount, $\Delta\lambda_1 = \lambda_o - \lambda_1$ and is obtained from the well-known formula

$$\lambda = \lambda_o\left(1 \pm \frac{v}{c}\right).$$

(λ_o is the input wavelength and λ is the wavelength observed by an observer moving with velocity v relative to the source of λ_o.)

This radiation is observed along the direction shown, and hence the observer will see radiation of wavelength λ_1 further Doppler-shifted by an amount depending on the component of velocity in the line of the detector. Let λ_2 be the wavelength of observed radiation.

Similarly, an electron at B with velocity equal to v^* will be observed to radiate at a wavelength λ_2 in the laboratory coordinates. We see, therefore, that all electrons whose velocity vector terminates on the extension of AB will be observed to give radiation shifted from the input wavelength by an amount $\Delta\lambda_2 = \lambda_o - \lambda_2$.

The total number of such electrons is simply the number lying in that slice of velocity space dV_a thick and V_a away from the origin and is given by

$$dN_{V_a} = Nf(V_a)\, dV_a,$$

where N is the total number of electrons, and f is the distribution function. For a Maxwell-Boltzmann distribution of electron velocities we have

$$dN_{V_a} = \frac{N}{\sqrt{\pi}}\left(\frac{m}{2kT}\right)^{1/2} \exp\left(-\frac{1}{2}\frac{mV_a^2}{kT}\right)dV_a,$$

which, in turn, can be written in terms of the wavelength shift $\Delta\lambda_2$ because

$$V_a^2 = \frac{V^{*2}}{2}(1+\cos\theta)$$

and

$$V^* = \frac{\Delta\lambda_2}{\lambda_o}c.$$

Hence

$$dN_{V_a} = \frac{N}{\sqrt{2\pi}}\left(\frac{m}{2kT_e}\right)^{1/2}\frac{c}{\lambda_o}(1+\cos\theta)^{1/2}\exp\left(-\frac{mc^2(1+\cos\theta)}{4k\lambda_o^2}\frac{\Delta\lambda_2^2}{T_e}\right)d\lambda,$$

where $d\lambda$ is the width of the filter used to select wavelength shift $\Delta\lambda_2$. For $\theta = 90°$ (the value used in our experiment) the expression simply reduces to

$$dN_{V_a} = A\exp\left(-\chi\frac{\Delta\lambda_2^2}{T_e}\right)d\lambda,$$

where

$$A = \frac{Nd\lambda}{\sqrt{2\pi}}\left(\frac{m}{2kT_e}\right)^{1/2}\frac{c}{\lambda_o}$$

and

$$\chi = \frac{mc^2}{4k\lambda_o^2}.$$

Hence a semi-log plot of signal amplitude versus $\Delta\lambda_2^2$ will give T_e for a Maxwell-Boltzmann distribution. The intensity of scattered radiation dI_s of wavelength λ_2 is

$$dI_s = \frac{I_o}{a}\left(\frac{e^2}{mc^2}\right)^2 \sin^2\phi\, d\Omega\, \frac{n_e V}{\sqrt{2\pi}}\left(\frac{m}{2kT_e}\right)^{1/2}\frac{c}{\lambda_o}\exp\left(-\chi\frac{\Delta\lambda_2^2}{T_e}\right)d\lambda$$

for plane polarized input, of intensity I_o in an area a. The angle ϕ is the angle between the plane of polarization and the direction of observation, and V is the volume in the field of view. Therefore, by making an absolute measurement of the intensity of scattered radiation, we can also determine the electron density n_e in the plasma.

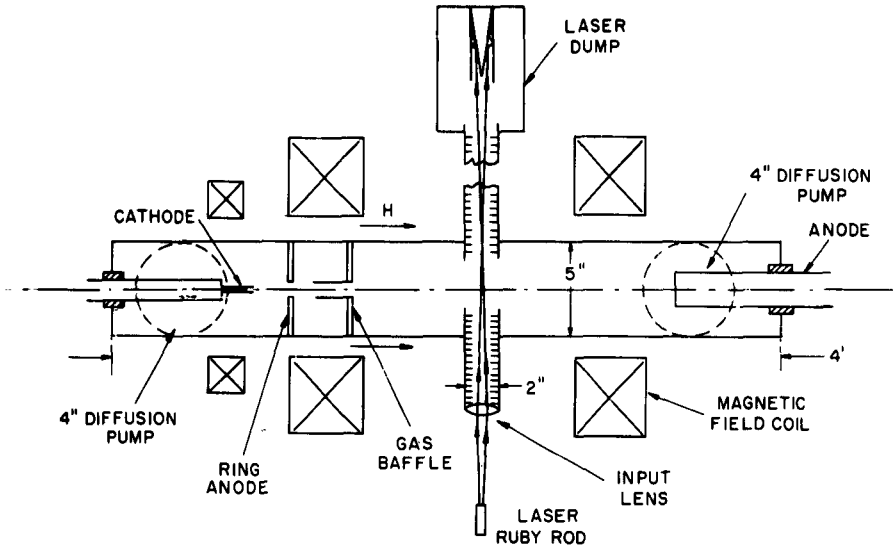


Fig. VII-20. Schematic diagram of apparatus. Plan view.

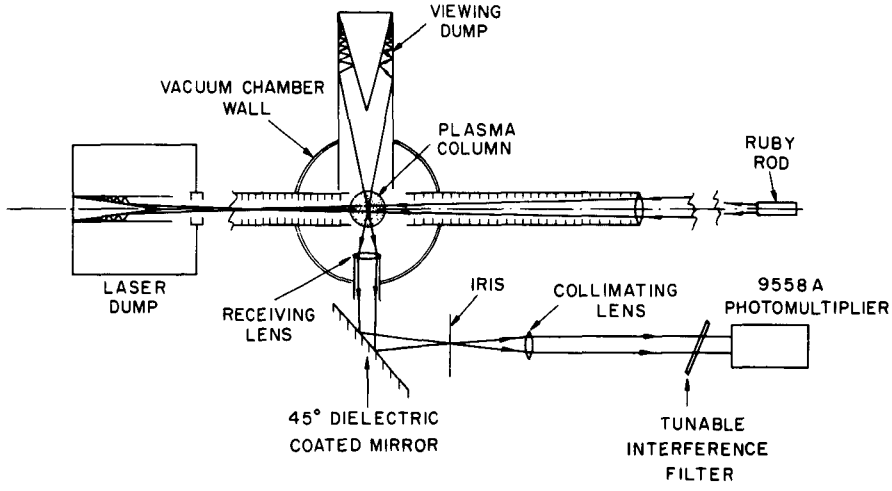


Fig. VII-21. Cross-section view of apparatus.

A schematic diagram of the apparatus is shown in Fig. VII-20 and a cross section through the plasma-laser interaction region is shown in Fig. VII-21. The plasma source is the hollow-cathode discharge developed in the Research Laboratory of Electronics, a full description of which has been given.[3] By using this device, an argon plasma with an electron density of $\sim 10^{13}$ electrons cm^{-3} and electron temperature of a few electron

volts is produced along the axis of the machine. Twenty joules of light at 6934 Å from a ruby laser is passed through the plasma column before being absorbed in the specially designed dump chamber, the principle of which is evident from the diagram. A similar dump forms a black background for the receiving lens whose field of view is limited by the iris. The scattered radiation is collimated before reaching the tunable interference filter that has a 5 Å passband, the center of which is continuously variable from 6940 Å to 6700 Å. The detector is an EMI 9558 A photomultiplier.

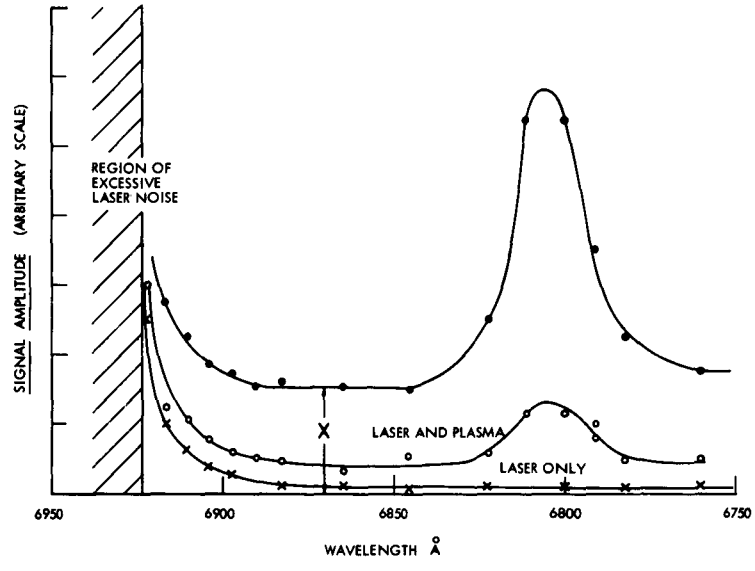


Fig. VII-22. Results of series 3 experiment (argon plasma stream).

The results of our first experiment (series 3), in which we obtained the spectrum of the scattered radiation, are given in Fig. VII-22 in which the amplitude of the detected signal (in arbitrary units) is plotted against wavelength. It is stressed, however, that these results are of a preliminary nature. The bottom curve is the spectrum obtained in the absence of plasma and represents the effect of radiation scattered from the walls of the vacuum chamber, and so forth, breaking through the filter. The middle curve shows the detector output when the laser is fired into the plasma, and the top curve represents the difference between these two curves on an expanded scale.

The shape of the top curve shows two peculiar effects:

(a) the peak centered at 6805 Å, and

(b) the curve does not go to zero in the wavelength range scanned.

It is difficult to attribute either of these effects to Thomson scattering from the

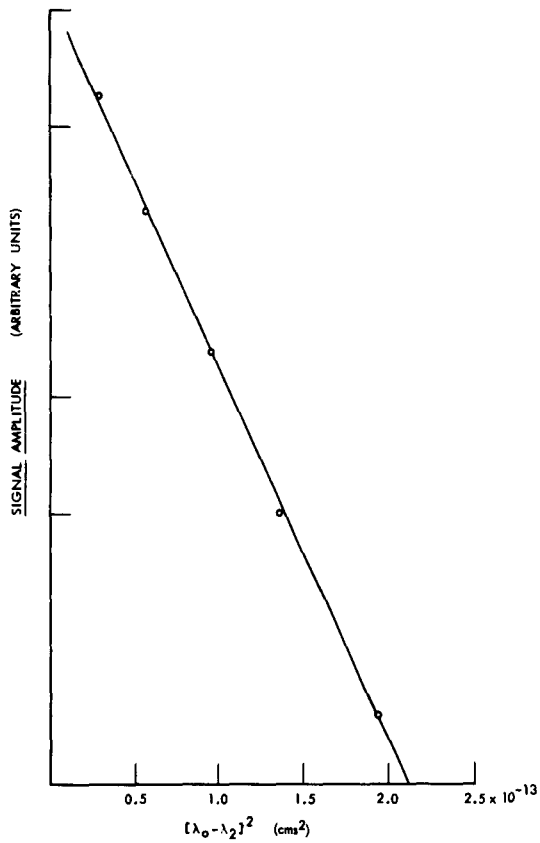


Fig. VII-23. Results of series 3 experiment.

plasma electrons, and it is tentatively suggested that these effects arise from the emission of radiation from the plasma as a result of the laser input. The large peak at 6805 Å is perhaps due to Raman scattering (the vibrational line being broadened by rotational lines).

We have some evidence to show that parts of the plasma emit radiation of fairly uniform intensity over a wide range of wavelengths ~0.2 msec after the laser pulse is turned on – this phenomenon suggests the idea of optical pumping of some levels in either excited states of A, A_2^+ or A^+ which then reradiate. If this unresolved effect of amplitude X is subtracted from the signal amplitude near the laser line, that is, we assume that there are no electrons present with energies greater than 10 ev, and plot signal amplitude versus $\Delta\lambda_2^2$, we obtain the result shown in Fig. VII-23. The value of electron temperature obtained from the gradient of this line is 1.2 ev.

E. Thompson, G. Fiocco

References

1. G. Fiocco and E. Thompson, Phys. Rev. Letters 10, 89 (1963).

2. G. Fiocco and E. Thompson, Scattering of light from electrons, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 90-91; and Quarterly Progress Report No. 68, pp. 74-76.

3. L. M. Lidsky, S. D. Rothleder, D. J. Rose, S. Yoshikawa, C. Michelson, and R. J. Mackin, Jr., J. Appl. Phys. 33, 2490-2497 (1962).

H. USE OF FISSILE NUCLIDES IN FUSION REACTOR BLANKETS

The preliminary investigation of this problem has been completed and a final report, in the form of an S.M. thesis to be submitted to the Department of Nuclear Engineering, M.I.T., is being prepared.[1]

The author's conclusions, based on tritium regeneration and heating-rate calculations, regarding the feasibility and the practicality of the three blanket configurations described previously[2,3] are as follows.

(a) The metallic thorium first-wall blanket assemblies are rejected on the ground that the thermal stress in the first wall limits the reactor operation to impractically low power levels.

(b) The blanket systems with depleted UF_4 fused salts in the first-wall coolant region are restricted to low-power operation (by heat transfer limitations arising from the liberation of from 77 per cent to 135 per cent of the incident neutron energy flux in the narrow coolant region).

(c) The blanket configurations containing a high UF_4 constituent (17-27 mole fraction percentage) fused salt in the primary attenuator with approximately 50 per cent Li^6 enrichment appear to be feasible and practical. The calculated tritium breeding ratios are in excess of 1.2; thus a self-sustaining tritium cycle is ensured. The heat-generation rates in the Mo first wall and in the coolant region are comparable to the nonfissile systems studied by Impink and Homeyer[4,5]; therefore operation at the same incident energy flux of from 4 to 5 $\frac{Mw}{m^2}$ is possible. The fast fission reaction doubles the recoverable energy liberated in the blanket, and approximately 90 per cent of the heat is generated in the large primary attenuator region. The manufacture of one plutonium 239 atom per five fusion reactions may have economic implications in terms of fission reactor fuel.

L. N. Lontai

References

1. L. N. Lontai, A Study of a Thermonuclear Reactor Blanket with Fissile Nuclides (S.M. Thesis, to be submitted to the Department of Nuclear Engineering, M.I.T., May 1963).

2. L. N. Lontai, D. J. Rose, and I. Kaplan, Use of fissile nuclides in fusion reactor blankets, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M.I.T., January 15, 1963, pp. 77-82.

3. L. N. Lontai and A. J. Impink, Jr., Use of fissile nuclides in fusion reactor blankets, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 91-94.

4. A. J. Impink, Jr., Neutron Economy in Fusion Reactor Blanket Assemblies, Ph.D. Thesis, Department of Nuclear Engineering, M.I.T., January 1963.

5. W. G. Homeyer, Thermal and Chemical Aspects of the Thermonuclear Blanket Problem, Sc.D. Thesis, Department of Nuclear Engineering, M.I.T., December 1962.

I. FUSION REACTOR BLANKET EXPERIMENT

1. Introduction

Experiments have been initiated to study the neutronics of models of a tritium regenerating blanket to surround a power-producing deuterium-tritium cycle thermonuclear reactor. The objects of the experiments are (a) to check calculational methods[1-3] used to determine neutron fluxes and tritium production rates, and (b) to provide a versatile

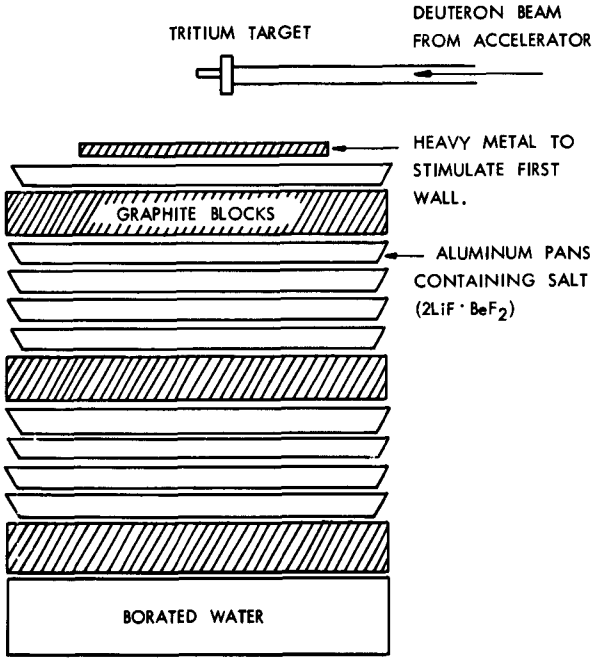


Fig. VII-24. Mock-up of a fusion reactor blanket.

means of studying effects of changes of materials and geometry on neutron flux and tritium production.

The blanket[2-5] is envisioned as a series of annular regions surrounding a cylindrical plasma, which contain materials for neutron multiplication and moderation, for removal of heat generated in the plasma and in the blanket, and for tritium production by the $Li^6(n, \alpha)H^3$ and $Li^7(n, n\alpha)H^3$ reactions. In the experiments, a core is removed from this blanket, and the resulting cylinder is irradiated with a point source of 14-Mev neutrons. A schematic diagram of the experimental arrangement is shown in Fig. VII-24.

2. Neutron Source

The 14-Mev neutrons for the experiment are supplied by a tritium gas target, shown in Fig. VII-25. The tritium is contained in a stainless-steel tube 1/4 inch ID and 1.3 inches long. The target window, a disc of commercial aluminum foil 33.8 μ thick, is

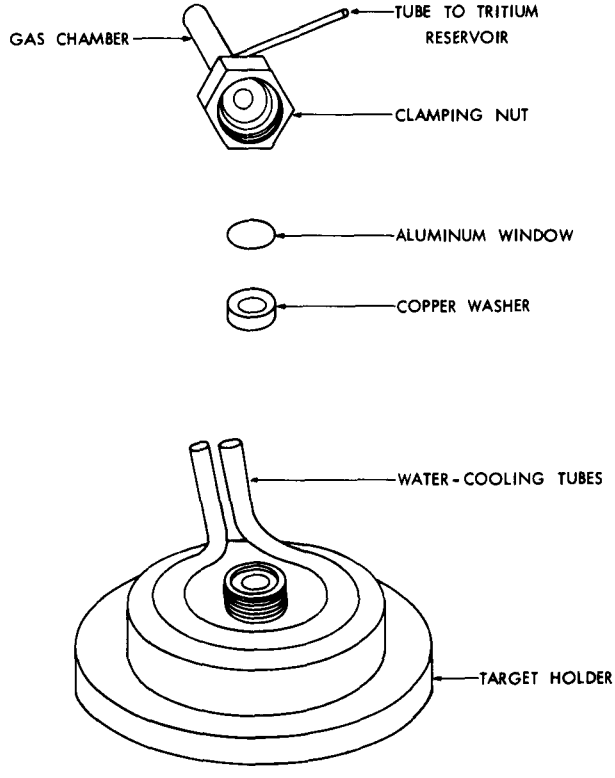


Fig. VII-25. Gas target.

held to the polished surface on a flange on the open end of the cylinder. A copper washer seals the window to the cylinder; the other side of the washer forms a seal with the target holder. Preliminary tests indicate that the window can withstand a 6-μa beam of 2.7-Mev deuterons almost indefinitely. The incident 2.7-Mev beam of deuterons is slowed to 400 kev in the window, whence it enters the gas chamber, where it is slowed to 40 kev. The calculated yield is 9×10^8 neutrons per second per microampere.

The tritium is supplied by a uranium tritide furnace similar to that described by Johnson and Banta.[6]

The 2.7-Mev deuterons are supplied by the Van de Graaff accelerator at the Air Force Cambridge Research Laboratory, Bedford, Massachusetts.

3. The Blanket Mock-Up

Most of the materials for the blanket, such as graphite for neutron-moderating regions, are available in sufficient quantities for the mock-up assembly. Approximately 300 lb of lithium-beryllium fluoride (composition $2LiF \cdot BeF_2$) to simulate the fused-salt regions have been supplied by the Reactor Chemistry Division of the Oak Ridge National Laboratory. The salt is sealed in 12 aluminum pans 18 inches in diameter, and 1 inch thick.

4. Tritium Counting

The tritium produced in the blanket assembly by the two lithium reactions is trapped in the crystal lattice of the lithium-beryllium fluoride. Salt samples will be melted in copper tube furnaces; the tritium thus evolved is swept into a proportional counter, by using a Teoppler pump to complete the transfer. A mixture of 10 per cent methane and 90 per cent argon is used as both the sweeping gas and the counting gas. After counting the tritium activity ($<10^5$ dpm) the gases in the proportional counter are exhausted outside the laboratory. Since tritium will build up in the bulk of the salt in the assembly from one experiment to another, fresh samples wrapped in aluminum foil will be used for counting in each experiment. By this method, the tritium production rate can be measured throughout the salt-containing regions of the assembly. The tritium-measuring system has been checked with salt samples irradiated by thermal neutrons in the M.I.T. reactor.

5. Threshold Detectors

a. Selection

The neutron flux in the blanket mock-ups will be measured as a function of position in the blanket and of the neutron energy by means of threshold detectors. After a search among publications five suitable threshold reactions were found: $U^{238}(n, f)$, $P^{31}(n, p)Si^{31}$, $Fe^{56}(n, p)Mn^{56}$, $I^{127}(n, 2n)I^{126}$, and $F^{19}(n, 2n)F^{18}$. Details of these detectors are given

Table VII-1. Threshold detector properties.

Element	Uranium	Phosphorus	Iron	Iodine	Fluorine
Reaction	$U^{238}(n, f)$	$P^{31}(n, p)Si^{31}$	$Fe^{56}(n, p)Mn^{56}$	$I^{127}(n, 2n)I^{126}$	$F^{19}(n, 2n)F^{18}$
Threshold Energy (Mev)	1.0	2.0	4.5	9.5	12.0
Half-life of daughter		2.62 hr	2.58 hr	2.6 hr (13 d)[a]	1.85 hr
"Effective" cross section[b]	990 mb	32.1 mb	34.9 mb	318.1 mb	21.0 mb
Material	Depleted metal foil	$Ca(PO_3)_2$ glass	Metal foil	PbI_2 or C_6I_6	Teflon (CF_2)
Competing reactions – Half-life of product		(n, 2n)-2.6 m (n, γ)-14.3 d	Fe^{54}(n, γ)-2.7 yr Fe^{58}(n, γ)-35 d	(n, γ)-25 m	(n, p)-29 s (n, α)-7.3 s (n, γ)-12 s
Radiation counted from main reaction	Integral γ	1.477 Mev β^-	845 kev γ	Integral β	650 kev β^+
References for cross-section curves	8	8, 9, 10	8, 11, 12, 13, 14, 15	8, 15, 16	16, 17, 18, 19, 20

[a] 2.6 hr is the internal transition half-life.

[b] Calculated for typical Impink spectrum.

in Table VII-1. Several factors had to be considered in choosing the reactions to be used.

Half-Life. Since it will require about one hour after a run with the accelerator to begin counting the foils, the half-life of the daughter products of the reactions that we use cannot be much less than 1 hour, or the activity produced will have decayed before it can be counted. On the other hand, if the half-life is much longer than several hours, not enough activity will be produced, unless the cross section for the reaction is very large. Also, the half-life of the daughters of the main reactions must be much longer or much shorter than those of competing reactions, so that either the competing activity can be allowed to decay before counting, or else the competing activity will be small.

Cross Section. The cross section for the reaction should be an accurately known function of energy in the region <16.5 Mev. This requirement eliminates most of the prospective reactions because (a) work with threshold detectors has usually been done for reactor systems, for which the highest energy of interest is ~2 Mev, and (b) a great deal of the data on threshold detectors results from measuring an effective cross section for a neutron energy spectrum that is almost a fission spectrum. From calculations made by Impink,[7] we conclude that the spectrum in the blanket assembly will be nothing

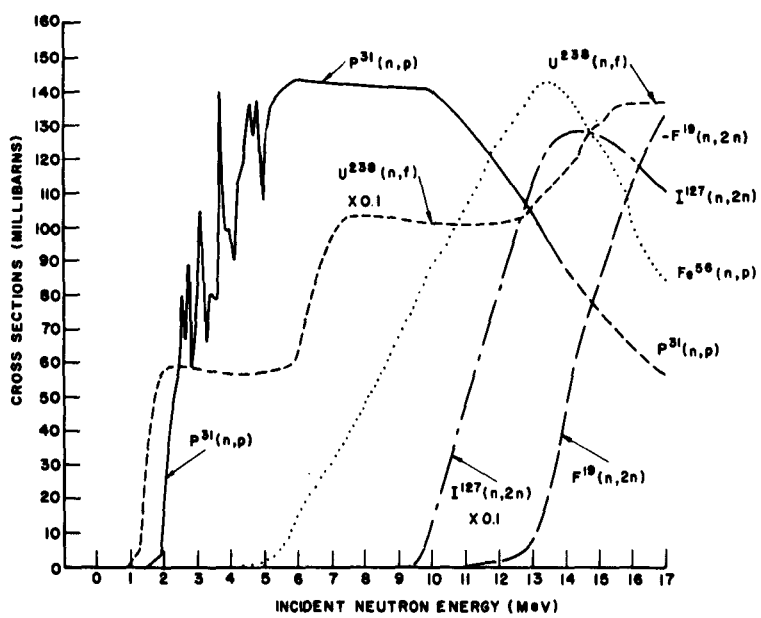


Fig. VII-26. Cross sections for threshold detector reactions. The curves shown are the best ones from available data (see Table VII-1). The curve for $P^{31}(n, p)$ has been averaged over the resonances in the 2-5-Mev region; beyond 14 Mev, it was extrapolated on a logarithmic plot.

like a fission spectrum; hence data on effective cross sections are of no value in the present work.

The cross sections for the reactions used are shown in Fig. VII-26. In general, the cross sections are those of BNL-325 modified by later measurements.[8] Except for the reaction $F^{19}(n, 2n)F^{18}$, the cross sections measured by various authors are in good agreement. In the case of the fluorine reaction there are two widely varying sets of measurements of $\sigma(E)$, one by French[17] researchers, and one by Russians.[18] The Russian data are rather carefully documented, but the French values are simply presented as preliminary results; furthermore, the Russian data at 14 Mev are in better agreement with the work of others who have measured the cross section only at this energy. Hence the Russian data are used in this work.

For use as a threshold detector the cross sections for the $P^{31}(n, p)Si^{31}$ reactions have been averaged over the many narrow resonances in the 2-5 Mev region. The cross-section curve for this reaction is extrapolated logarithmically above 14 Mev, because of the lack of measurements in this region. The extrapolation seems reasonable from other (n, p) reactions that were studied.

Foil Material. To ease the problem of counting the foils after a given run, the parent isotope must be in a form that does not require special handling procedures. If a compound is to be used, the additional elements cannot contribute too much to the background. For uranium and iron, metal foils can be used to satisfy this requirement. The carbon in the Teflon that is used for the fluorine does not interfere, since the threshold energy for the $C^{12}(n, 2n)$ reaction (20 Mev) is above the maximum neutron energy that will be encountered in these experiments.

Although most authors suggest using ammonium phosphate as a phosphorous-containing foil, we find that calcium metaphosphate ($Ca(PO_3)_2$) is a better material.[21] The foils were manufactured as follows: Monobasic calcium phosphate powder ($Ca(H_2PO_4)_2 \cdot H_2O$) is heated in graphite crucibles to 2300°F, at which temperature it dehydrates, decomposes to $Ca(PO_3)_2$, and then melts. The resulting clear syrupy fluid is poured into graphite molds (which are heated by a hot plate), and pressed into shape, before molten fluid solidifies. The resulting clear glass pellets (3/4 inch in diameter, 1/8 inch thick) are annealed at 850°F for several hours to remove the thermal strains introduced in casting, and then slowly brought to room temperature. The pellets can then be handled as ordinary glass, although they are not quite so brittle. They are of approximately 90 per cent theoretical density, because of air bubbles and bits of graphite trapped during the casting process.

For the $I^{127}(n, 2n)I^{126}$ reaction, lead iodide[22] appears to be a suitable material. Hexaiodobenzene is also being investigated.

b. Analysis of Data

The analysis of the threshold foil data is based on an orthonormal polynomial expansion method of Lanning and Brown.[23] A FORTRAN II program has been written to perform the data analysis, and is now being tested with reaction rates calculated by Impink[7] used.

P. S. Spangler

References

1. W. G. Homeyer, A. J. Impink, Jr., D. J. Rose, and I. Kaplan, The fusion blanket problem, Quarterly Progress Report No. 62, Research Laboratory of Electronics, M.I.T., July 15, 1961, pp. 64-68.

2. W. G. Homeyer and A. J. Impink, Jr., Energy extraction blanket for a fusion reactor, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M.I.T., January 15, 1962, pp. 128-131.

3. W. G. Homeyer, A. J. Impink, Jr., and D. J. Rose, Energy extraction blanket for a fusion reactor, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 142-150.

4. A. J. Impink, Neutron Economy in Fusion Reactor Blanket Assemblies, Sc.D. Thesis, Department of Nuclear Engineering, M.I.T., January 1963.

5. D. J. Rose and M. Clark, Jr., Plasmas and Controlled Fusion (The M.I.T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961).

6. C. H. Johnson and H. E. Banta, Rev. Sci. Instr. 27, 132 (1956).

7. A. J. Impink, Jr., Private communication, October 1962.

8. BNL-325, Brookhaven National Laboratory, Upton, New York, Second Edition, 1958, Supplement, 1960.

9. P. Cuzzorrea, G. Pappalardo, and R. Ricamo, Nuovo cimento 16, 450 (1960).

10. J. Kantele and D. G. Gardner, Nuclear Phys. 35, 353 (1962).

11. J. Depraz, G. Legros, and R. Salin, J. phys. radium 21, 377 (1960).

12. S. Yasumi, J. Phys. Soc. Japan 12, 443 (1957).

13. B. D. Kern, W. E. Thompson, and J. M. Ferguson, Nuclear Phys. 10, 233 (1959).

14. D. M. Chittenden II, D. G. Gardner, and R. W. Fink, Phys. Rev. 122, 860 (1961).

15. M. Bormann, S. Cierjacks, R. Lang Kau, and H. Neuert, Z. Phys. 166, 477 (1962).

16. E. B. Paul and R. L. Clark, Can. J. Phys. 31, 267 (1953).

17. C. W. Williamson and J. Picard, J. phys. radium 22, 651 (1961).

18. O. D. Brill', N. A. Vlasov, S. P. Katinin, and S. L. Sokolov, Doklady Akad. Nauk S.S.S.R. 136, 55 (1961).

19. L. A. Rayburn, Bull. Am. Phys. Soc., Ser. II 3, 377 (1958).

20. V. J. Ashby, H. C. Catron, L. L. Newkink, and C. J. Taylor, Phys. Rev. 111, 616 (1958).

21. L. H. Weinberg, R. P. Schuman, and B. A. Gottschalk, KAPL-1283 Knolls Atomic Power Laboratory, 1955.

22. V. W. Brideweser, S.M. Thesis, Department of Nuclear Engineering, M.I.T., June 1961.

23. W. D. Lanning and K. W. Brown, WAPD-T-1380, Bettis Atomic Power Laboratory, Pittsburg, Pennsylvania, 1961; see also Trans. Am. Nuclear Soc. 4, 267(1961).

VIII. PLASMA MAGNETOHYDRODYNAMICS AND ENERGY CONVERSION*

Prof. G. A. Brown
Prof. E. N. Carabateas
Prof. S. I. Freedman
Prof. W. H. Heiser
Prof. M. A. Hoffman
Prof. W. D. Jackson
Prof. J. L. Kerrebrock
Prof. J. R. Melcher
Prof. J. P. Penhune
Prof. A. H. Shapiro
Prof. R. E. Stickney
Prof. H. H. Woodson
M. T. Badrawi
A. N. Chandra
R. S. Cooper
J. M. Crowley
D. A. East
J. R. Ellis, Jr.
J. Gerstmann
N. Gothard
J. B. Heywood
H. D. Jordan
P. G. Katona
F. D. Ketterer
G. B. Kliman
H. C. Koons
M. F. Koskinen
R. F. Lercari
W. H. Levison
A. T. Lewis
H. C. McClees, Jr.
C. W. Marble
T. D. Masek
J. T. Musselwhite
S. A. Okereke
J. H. Olsen
C. R. Phipps, Jr.
E. S. Pierson
D. H. Pruslin
M. H. Reid
C. W. Rook, Jr.
A. W. Rowe
M. R. Sarraquigne
M. Shroff
P. M. Spira
R. Toschi
G. L. Wilson
J. C. Wissmiller
B. M. Zuckerman

A. THERMIONIC EMISSION FROM MONOCRYSTALLINE TUNGSTEN SURFACES IN THE PRESENCE OF CESIUM

1. Introduction

The preliminary results of a recently initiated thermionic investigation are reported here, together with a brief description of our apparatus. The primary purpose of this study is to attain a better fundamental understanding of the emission processes that occur at the electrodes of thermionic energy converters.

A well-defined emitter surface is necessary in order to obtain experimental results of the quality that is most useful in the formulation of a general theory of thermionic processes. Monocrystalline surfaces are generally accepted as being the best specimens for such studies, and many investigators have employed them with varying degrees of success. Two of the most significant experiments are those of Nichols[1] and Smith,[2] the results of which show that the work function of clean tungsten depends markedly on the crystallographic structure of the particular surface that is being studied. The close-packed surfaces generally have the higher work functions, the approximate range for tungsten being from 4.65 ev down to 4.30 ev.

Although much work has been devoted to the study of clean monycrystalline surfaces, there have been very few quantitative investigations of monocrystalline surfaces that are partially covered with a layer of adsorbed molecules, a problem of current importance to thermionic energy conversion and ion propulsion. Qualitative observations have been made by a number of investigators,[3,4] and Webster[5] has obtained significant

*This work was supported in part by the National Science Foundation under Grant G-24073, and in part by the U.S. Air Force (Aeronautical Systems Division) under Contract AF33(616)-7624 with the Aeronautical Accessories Laboratory, Wright-Patterson Air Force Base, Ohio.

quantitative results for a few of the most conspicuous crystallographic directions of several refractory metals. The results show that cesium affects the close-packed planes most strongly, causing their work functions to be reduced by as much as 3 ev.

There is a definite need for additional research in this area since the existing theories are incomplete, and the experimental results are insufficient. For this reason, the present investigation has been designed to study in detail the dependence of the thermionic properties of tungsten on crystallographic direction, surface temperature, and cesium density.

2. Apparatus

An experimental tube built by Salomon[6] was modified for the present investigation. The design is basically a simplification of that used by Nichols.[1] The test specimen is a tungsten filament having a diameter of 0.0049 inch and a length of 14 cm. A long crystal has been grown in this filament by means of a special heating schedule.[4] Crystals formed in this manner are usually oriented in such a manner that the (110) crystallographic direction is parallel to the axis of the wire; this means that all crystallographic directions that are normal to the surface of the filament will have Miller indices of the form (hhk). Martin[3] has shown that this class of directions includes those of maximum and minimum emission; therefore it is most desirable for thermionic studies.

The filament is surrounded by a three-piece cylindrical anode (Fig. VIII-1). The two endpieces are fixed and serve as guard rings, while the center section is rotatable (by means of magnets) and contains a small slit cut parallel to the axis of the filament. The diameter of the center section is 1.9 cm, and the slit dimensions are 0.07 cm by 1.01 cm, so that the subtended angle is approximately 4.23°. By applying a strong electric field between the filament and the three-piece anode, one can cause the electrons emitted from the surface to be accelerated radially in straight-line paths. The slit acts as a microscope focused on a small area of the emitting surface, thereby allowing only the electrons from that area to pass through the slit to be measured at the collector. (The approximate value of this area for the present tube is 4.65×10^{-4} cm^2 if the filament-anode concentricity is perfect.) It is therefore possible, at least in principle, to measure the emission associated with the different crystallographic directions by rotating the slit to the appropriate positions.

A schematic of the electrical circuit is included in Fig. VIII-1. A dc power supply is used to maintain the three-piece anode at a potential of +1000 volts with respect to the filament. The cylindrical collector is held at potential of only 10 volts above that of the filament to reduce the undesirable effects of secondary emission. The collector current is measured with a suitable electrometer.

A cesium reservoir is connected to the experimental tube, and its temperature is regulated in order to attain the desired density of the cesium vapor. The tungsten

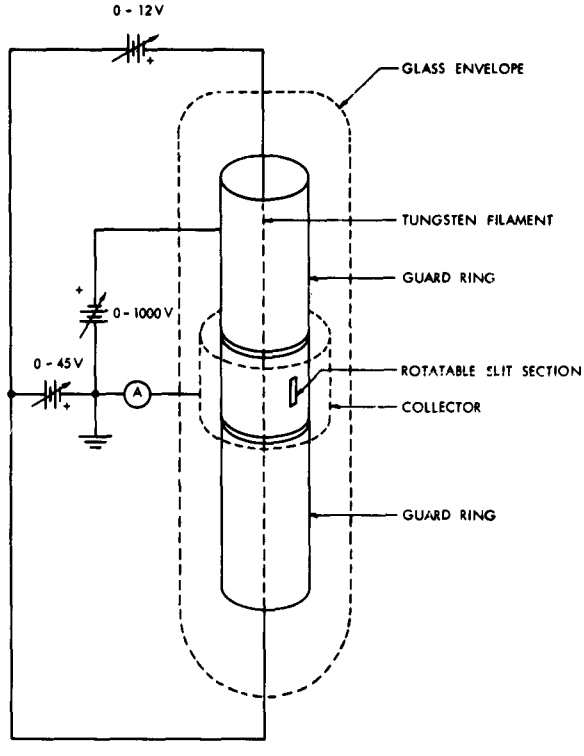


Fig. VIII-1. Experimental apparatus.

filament is heated directly by ac current, and its temperature is determined simply from the resistivity, a method that is of questionable accuracy at temperatures less than 1300°K unless the appropriate calibrations and corrections are made.[1]

3. Results

The tube was evacuated by means of a mercury diffusion-pump system, and a pressure of approximately 10^{-10} mm Hg was attained after thorough out-gassing. The filament was flashed a number of times until the emission pattern appeared to be stable. The filament temperature was then set at 1610°K, and readings of the collector current were taken for various slit positions. A portion of the data is shown in Fig. VIII-2, and its general pattern is similar to those reported by Nichols[1] and Smith.[2] Therefore, by referring to their work it was possible to estimate the positions of the prominent crystallographic directions along the abscissa. The abscissa actually represents the angular position of the slit; the pattern in Fig. VIII-2 does not cover the full 360° of azimuth, as its scope was limited to approximately 210°.

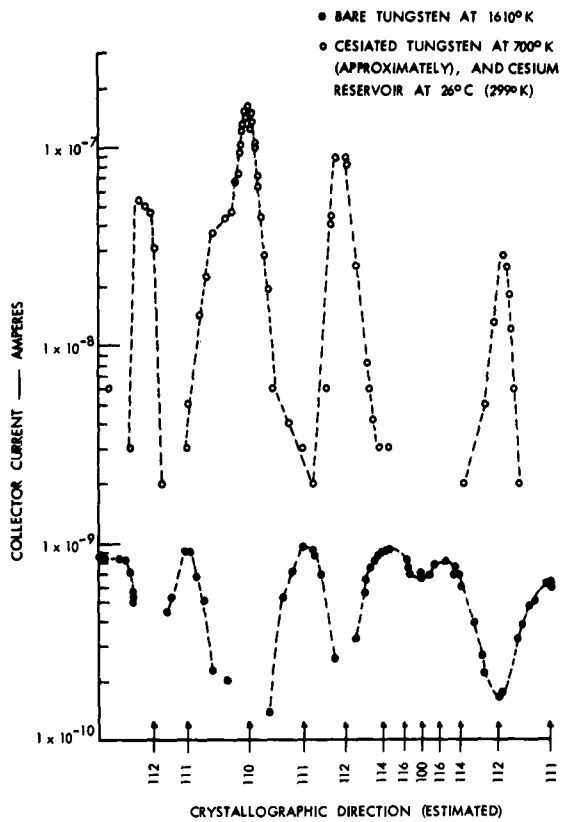


Fig. VIII-2. Thermionic emission vs crystallographic direction for cesiated and bare tungsten.

The tube was then sealed off from the vacuum system, and cesium was introduced by the usual method, that is, breaking a cesium-filled glass capsule in the reservoir. With the cesium reservoir at room temperature (26°C), the filament temperature was so adjusted as to obtain maximum emission in the (110) direction. (This temperature is not known accurately because a detailed calibration has not yet been performed; the estimated value is 700°K.) The resulting emission data for various slit positions are included in Fig. VIII-2 for comparison with the previous data obtained for bare tungsten. Since these are preliminary results, a detailed analysis is not warranted at this time. However, it is interesting to note that:

(i) The presence of cesium causes the maximum emission at 700°K to be as much as two orders of magnitude greater than that for the bare surface at 1610°K.

(ii) Surfaces having the lowest work function in the presence of cesium are those that have the highest "bare" work function.

Both of these general observations are in agreement with the results obtained by

other investigators.[3-5] On the basis of these preliminary results, we conclude that the present technique may be employed now in a more detailed study of emission processes.

4. Future Plans

The filament failed before more reproducible data could be obtained. A new filament is being prepared, and modifications have been made to reduce the effects of stray currents caused by leakage, secondary electrons, and photoemission. The immediate plan is to obtain data for other conditions of emitter temperature and cesium density.

The close cooperation of the members of the Physical Electronics Group of the Research Laboratory of Electronics has been invaluable throughout this program.

R. E. Stickney

References

1. M. H. Nichols, The thermionic constants of tungsten as a function of crystallographic direction, Phys. Rev. 57, 297 (1940).

2. G. F. Smith, Thermionic and surface properties of tungsten crystals, Phys. Rev. 94, 295 (1954).

3. S. T. Martin, On the thermionic and adsorptive properties of the surfaces of a tungsten single crystal, Phys. Rev. 56, 947 (1939).

4. G. G. Zipfel, S.B. Thesis, Department of Mechanical Engineering, M.I.T., 1960.

5. H. F. Webster, Thermionic emission from a tantalum crystal in cesium or rubidium vapor, J. Appl. Phys. 32, 1802 (1961).

6. S. N. Salomon, S.B. Thesis, Department of Mechanical Engineering, M.I.T., 1961.

B. A-C POWER GENERATION WITH MAGNETOHYDRODYNAMIC CONDUCTION MACHINES

Magnetohydrodynamic (MHD) power generation is attractive because of the inherent simplicity of the system. When an ionized gas is used as the working fluid, higher temperature operation is possible with an MHD generator. This makes heat rejection easier in space applications and increases thermal efficiency in ground-based installations.

In many applications, especially those in which the power is to be processed, it is desirable to generate ac power. Direct generation of ac power in MHD machines with ionized gases used as the working fluids has been studied extensively.[1] The several theoretical possibilities proposed thus far have been shown to be impractical for operation with conditions that are characteristic of combustion temperatures, either because electrical conductivity of the gas is too low or because too much external capacitance is necessary to provide the reactive power to excite the magnetic field.[1]

In this report we describe a new type of ac MHD power generator. It is a conduction

machine, and all reactive power is handled by inductive energy storages. Moreover, the criterion for successful operation as a self-excited ac machine is simply that the machine be suitable for self-excitation as a dc machine.

Although the system is most promising for application to gaseous MHD systems, the fundamental mechanism of operation will be demonstrated theoretically with an incompressible fluid mode. This is standard practice with many types of MHD conversion devices. After the system has been analyzed with scalar conductivity assumed, a configuration will be presented which is suitable for gases exhibiting a large Hall effect. Finally, some estimates of performance will be made for gas properties that are characteristic of temperatures at which gases can be contained by material walls.

1. System with Scalar Conductivity

Consider the system illustrated in Fig. VIII-3. It consists of a channel of constant cross-section area having the dimensions shown. An incompressible fluid flows with the velocity $\overline{v}$ in the direction shown. There are two pairs of electrodes feeding, through two sets of coils, two identical load resistances R_L. The generators formed by the two pairs of electrodes are assumed to be independent insofar as the fluid is concerned. That is, end and edge effects are neglected and the two circuits are electrically insulated

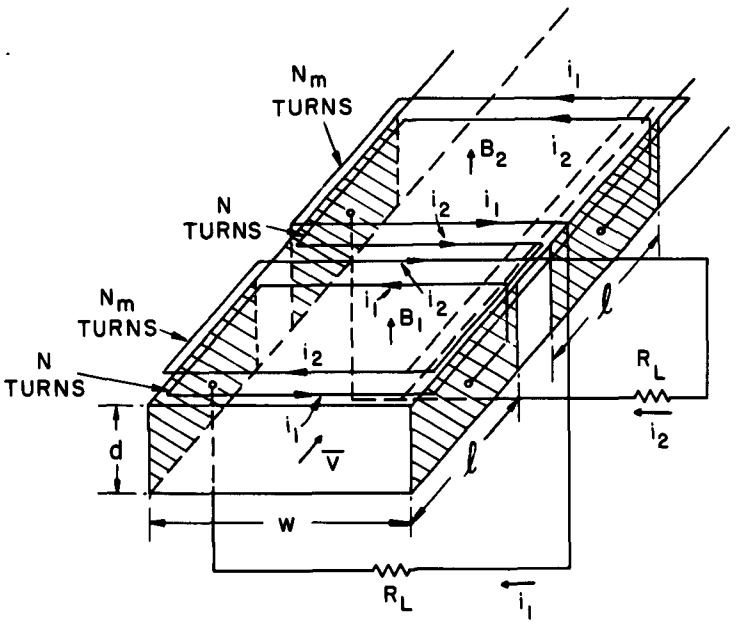


Fig. VIII-3. Basic configuration with solid electrodes.

from each other. Each of the two generators has two magnet coils, one carrying current i_1 and the other carrying current i_2 in the directions indicated. In practice, each coil would have a second half under the channel with the two connected series-aiding. Only one-half of each coil is shown in Fig. VIII-3 for simplicity. Although only two sets of electrodes and coils are shown, the system can have any number of sets. Furthermore, the sets can be connected for two-phase operation as in Fig. VIII-3 or for operation with any number of phases. In the last case, the number of coils associated with each pair of electrodes must be increased, but the essential features of ac operation are the same as those described below.

Denoting the magnetic flux density B_1 and B_2 in the regions of the two electrodes, assuming the positive directions shown in Fig. VIII-3, assuming the coils to be distributed so that the flux densities are essentially uniform over the length ℓ, and assuming that the coils produce a field just as a long solenoid (or assuming infinitely permeable pole pieces), we can write

$$B_1 = \frac{N\mu_o}{d} i_1 - \frac{N_m\mu_o}{d} i_2 \tag{1}$$

$$B_2 = \frac{N\mu_o}{d} i_2 + \frac{N_m\mu_o}{d} i_1 . \tag{2}$$

If we define the internal resistance for one pair of electrodes as

$$R_i = \frac{w}{\ell d\sigma} , \tag{3}$$

where σ is the electrical conductivity of the fluid, we can write the equations for the two circuits as

$$vwB_1 = Ri_1 + N\ell w \frac{dB_1}{dt} + N_m \ell w \frac{dB_2}{dt} \tag{4}$$

$$vwB_2 = Ri_2 + N\ell w \frac{dB_2}{dt} - N_m \ell w \frac{dB_1}{dt} . \tag{5}$$

Here, R is the total series resistance in one circuit ($R = R_L + R_i + R_c$, where R_c is the resistance of the magnet coils).

Substituting Eqs. 1 and 2 in Eqs. 4 and 5 and simplifying yields

$$Gi_1 - G_m i_2 = Ri_1 + L \frac{di_1}{dt} \tag{6}$$

$$Gi_2 + G_m i_1 = Ri_2 + L \frac{di_2}{dt} , \tag{7}$$

in which the parameters G, G_m, and L are defined as

$$G = \frac{vw\mu_o N}{d} \tag{8}$$

$$G_m = \frac{vw\mu_o N_m}{d} \tag{9}$$

$$L = \frac{\mu_o \ell w}{d}\left(N^2+N_m^2\right). \tag{10}$$

Elimination of i_2 from Eqs. 6 and 7, after some simplification, yields

$$\frac{d^2 i_1}{dt^2} + \frac{2(R-G)}{L}\frac{di_1}{dt} + \left[\frac{G_m^2 + (R-G)^2}{L^2}\right] i_1 = 0. \tag{11}$$

This equation has solutions of the form

$$i_1 = I_o \exp\left(-\frac{(R-G)}{L}t\right)\cos\left(\frac{G_m}{L}t + \theta\right), \tag{12}$$

where I_o and θ are constants to be evaluated from initial conditions. It is clear from the form of this solution that when $R > G$ the response is a damped sinusoid, when $R < G$ the response is an exponentially growing sinusoid, and when

$$R = G \tag{13}$$

the response is a sinusoid of constant amplitude. Reference to Eqs. 6 and 7 shows that Eq. 13 is the condition for self-excited operation as a dc generator.

Henceforth we shall assume that Eq. 13 is satisfied, in which case Eqs. 6 and 7 can be written

$$-G_m i_2 = L\frac{di_1}{dt} \tag{14}$$

$$G_m i_1 = L\frac{di_2}{dt}. \tag{15}$$

These equations, used with Eq. 12, show that i_1 and i_2 form a two-phase set. That is, if

$$i_1 = I_o \cos(\omega_m t+\theta), \tag{16}$$

Eq. 14 shows that

$$i_2 = I_o \sin(\omega_m t+\theta). \tag{17}$$

With the system parameters adjusted to satisfy Eq. 13, the system operates in the steady state with frequency

$$\omega = \frac{G_m}{L} \tag{18}$$

and with currents of finite amplitude. Consequently, power is dissipated in R_L and useful ac power is generated.

Equation 13 is the condition that must be satisfied for self-excitation as a dc generator with N turns on the field winding, and a coil resistance R_c in series with the load. The addition of the coupling windings N_m has transformed this pair of self-excited dc generators into a two-phase, self-excited ac generator. The price for making the change to ac operation is the addition of the N_m-turn coils that occupy space and contribute to resistance R_c. Reactive power is handled completely by the two magnetic storages having flux densities B_1 and B_2. Note that the two currents have 90° phase difference. The use of these currents in Eqs. 1 and 2 shows that B_1 and B_2 also have 90° phase difference. Thus the magnetic energy is constant in magnitude and oscillates between the two magnetic-field storages.

2. Limiting Parameters

Assuming steady-state ac operation, we can use the one-dimensional model in the analysis above to identify some limiting parameters of the system.

The use of Eqs. 9 and 10 in Eq. 18 yields

$$\omega = \frac{vN_m}{\ell\left(N^2+N_m^2\right)}. \tag{19}$$

It will be desirable to operate such a machine at as high a frequency as possible. With fixed v and ℓ, Eq. 19 has a maximum value when $N_m = N$ (we assume here, of course, that a minimum of one turn is possible). Thus, we can define the maximum frequency of operation as

$$\omega_m = \frac{v}{2\ell N} \tag{20}$$

with the constraint that $N = N_m$.

Next, we can assume that $R_L + R_c$ is some multiple of R_i; thus

$$R_L + R_c = \alpha R_i. \tag{21}$$

Then, using the definition of R_i (Eq. 3) and the condition for self-excitation (Eq. 13), we obtain

$$N = \frac{1+\alpha}{\mu_o \sigma \ell v}. \tag{22}$$

Note that the denominator of this expression is a magnetic Reynolds number based on

the length of a pair of electrodes.

Substitution of Eq. 22 in Eq. 20 yields

$$\omega_m = \frac{\mu_o \sigma v^2}{2(1+\alpha)}. \tag{23}$$

This result shows that the maximum frequency of operation depends on the conductivity σ, velocity v, and degree of loading α, and is independent of the size of the device.

To obtain some idea of the conditions that must be satisfied with an ionized gas for ac power generation of this type, we can use Eq. 23 to calculate values. Assuming a frequency of 60 cps, we can rewrite Eq. 23 as

$$\frac{\sigma v^2}{1+\alpha} = 6 \times 10^8.$$

Values of α, σ, and v that satisfy this equation are given in Table VIII-I.

Table VIII-I. Velocity, conductivity, and loading factor for operation at 60 cps.

loading factor α	0	0.5	1.0	2.0
(v m/sec)		(σ mhos/meter)		
1000	600	900	1200	1800
1500	270	400	530	800
2000	150	230	300	450
2500	96	144	190	290
3000	67	100	130	200
3500	49	74	100	150
4000	37	55	75	110

It is apparent from the numbers in Table VIII-I that there is a chance of using this type of generator with combustion gases and with gases having increased conductivity that is due to nonequilibrium effects. At a given velocity and loading factor the required conductivity varies linearly with frequency. Thus the device will work better at frequencies below 60 cps. For gases at combustion temperatures, high-frequency operation (such as 400-1000 cps) does not appear possible because of the high velocities and conductivities required.

The data of Table VIII-I can be used with an assumed length in Eq. 22 to find the number of turns required. To show that the number of turns is reasonable, assume ℓ = 1 meter and use α = 1, v = 3000 m/sec, σ = 130 mhos/meter from Table VIII-I to find N = 4 turns; this value is certainly reasonable for constructing such a device. The

other conditions of Table VIII-I will yield numbers of turns having the same order of magnitude.

3. Segmented Electrodes and Tensor Conductivity

It appears from the numbers mentioned in the preceding section that the number of turns per field coil may be small. For a large generator, the field conductors in such a system may be so large that eddy currents become a problem. Moreover, if the gas has tensor conductivity, the solid electrodes will short out the Hall electric fields and degrade performance.

Both of these deleterious effects can be reduced by using segmented electrodes. The construction for two electrode segments per generator section is illustrated in Fig. VIII-4. The essence of this arrangement is that each electrode pair is connected to a circuit exactly like that for solid electrodes in Fig. VIII-3. The circuit for each

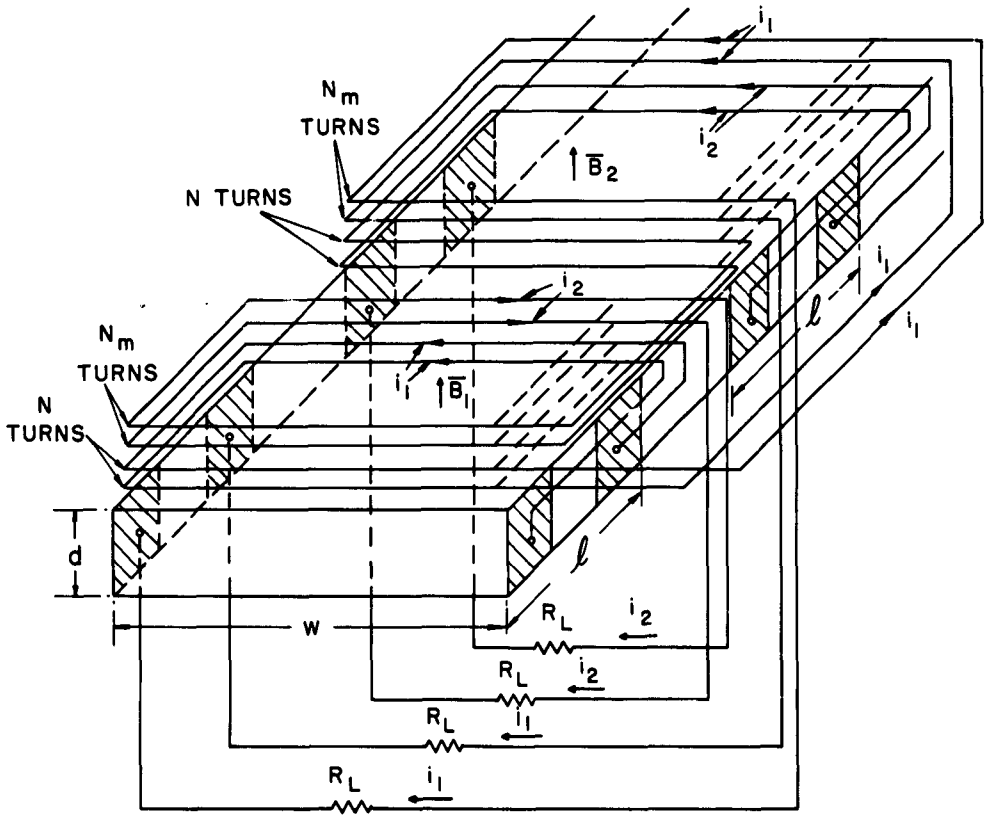


Fig. VIII-4. Configuration for use with segmented electrodes.

electrode pair is electrically insulated from all other circuits. If many electrode pairs are used, the power will probably be extracted by inductive coupling to magnetic fields B_1 and B_2, or by replacing R_L with a transformer winding. In any case, the generation of ac power with segmented electrodes does not create the problem of having many insulated loads that a dc generator does.

For the analysis assume that there are n electrode segments per generator (of length ℓ) and that the coils span a distance ℓ in the flow direction. If each pair of electrodes is assumed to carry the same current as other segments in the same generator section, then for the flux densities B_1 and B_2 (see Eqs. 1 and 2) we can write

$$B_1 = \frac{nN\mu_o}{d} i_1 - \frac{nN_m\mu_o}{d} i_2 \tag{24}$$

$$B_2 = \frac{nN\mu_o}{d} i_2 + \frac{nN_m\mu_o}{d} i_1. \tag{25}$$

Assuming a one-dimensional model, the internal resistance R_i' associated with a pair of electrodes (see Eq. 3) is

$$R_i' = nR_i = \frac{nw}{\ell d\sigma}. \tag{26}$$

The equations for single electrode pairs in the two generator sections (see Eqs. 4 and 5) are

$$vwB_1 = R'i_1 + N\ell w \frac{dB_1}{dt} + N_m \ell w \frac{dB_2}{dt} \tag{27}$$

$$vwB_2 = R'i_2 + N\ell w \frac{dB_2}{dt} - N_m \ell w \frac{dB_1}{dt}, \tag{28}$$

where $R' = R_L + R_i' + R_c$ is the total series resistance in one electrode-pair circuit.

Substituting Eqs. 24 and 25 in Eqs. 27 and 28 and simplifying (see Eqs. 6 and 7) yields

$$G_n i_1 - G_{mn} i_2 = R'i_1 + L_n \frac{di_1}{dt} \tag{29}$$

$$G_n i_2 + G_{mn} i_1 = R'i_2 + L_n \frac{di_2}{dt}. \tag{30}$$

The parameters G_n, G_{mn}, and L_n are defined as

$$G_n = \frac{nN\mu_o vw}{d} \tag{31}$$

$$G_{mn} = \frac{nN_m\mu_o vw}{d} \tag{32}$$

$$L_n = \frac{n\mu_o \ell w}{d}\left(N^2+N_m^2\right) \tag{33}$$

For this system self-excitation is achieved when the parameters are adjusted to satisfy the condition

$$G_n = R'. \tag{34}$$

With this condition satisfied, the frequency of steady-state operation is

$$\omega = \frac{G_{mn}}{L_n}. \tag{35}$$

By using the definitions of Eqs. 32 and 33, the frequency can be expressed as

$$\omega = \frac{vN_m}{\ell\left(N^2+N_m^2\right)}. \tag{36}$$

This is exactly the same as Eq. 19 for the system with solid electrodes.

The maximum frequency occurs when $N_m = N$, and is

$$\omega_m = \frac{v}{2\ell N}. \tag{37}$$

Defining the parameter a as

$$R_L + R_c = aR_i' \tag{38}$$

we can use the self-excitation condition, Eq. 34, with Eq. 26 to obtain

$$N = \frac{1+a}{\mu_o \sigma \ell v}. \tag{39}$$

Note that the results of Eqs. 36, 37, and 39 are independent of the number of electrode segments and are the same as Eqs. 19, 20, and 22 for the system with solid electrodes. Thus, all of the discussion of limiting parameters in the preceding section holds also for the case of segmented electrodes. The only effect of segmenting the electrodes is to require a number of windings that is equal to the number of electrodes, each winding having the same number of turns as required by the machine with solid electrodes. Each winding carries a fraction 1/n of the total current supplied by a generator section; consequently, the volume of field conductor is independent of the number of electrode segments.

4. Discussion

The generators discussed in this report on an idealized basis indicate the possibility of their being applicable to systems in which ionized gases are used at temperatures

that are characteristic of combustion processes. Their advantage is that they generate alternating current directly and use only magnetic fields for energy storage.

The theory assumes incompressible flow. The basic mechanism still occurs in compressible flow, the velocity being the average flow velocity of the gas. There will be compressibility effects that need to be evaluated and they may even augment the conversion process.

The generators analyzed here are two-phase machines. By a simple extension of the concepts we can devise an n-phase machine by having n separate sets of electrodes and n sets of magnet coils along the channel. A coupling coil is required in each set for current from every other set of electrodes. No matter how many phases a machine may have, it can be reduced to an equivalent two-phase machine for analysis. Thus the present analysis is complete for a balanced n-phase machine.

H. H. Woodson

References

1. H. H. Woodson, Magnetohydrodynamic A-C Power Generation, Proc. AIEE Pacific Energy Conversion Conference, San Francisco, California, August 1962.

C. ELECTROHYDRODYNAMIC WAVES IN ROTATIONAL SYSTEMS

1. Introduction

Efforts to achieve a stable magnetohydrodynamic containment have led some investigators to consider the effect of rotation. From everyday experience, one would suspect that rotational acceleration, just as the acceleration of gravity, could be responsible for stabilizing or unstabilizing a system, the choice depending on the configuration. Common examples are the gravitational or Taylor instability, produced by gravity on a top-heavy fluid interface[1] or the stabilizing influence of gravity in a situation in which a perpendicular electric field is responsible for a surface instability.[2]

However, it is well known that rotational motion is considerably more involved than this. In addition to the centrifugal forces experienced in pure rotation, particles that are in motion with respect to a rotating frame respond to a Coriolis acceleration, which results in a mode splitting that has been well known since the last century in conjunction with ordinary gravity waves,[3] but has been studied in detail only recently.[4]

This report is a study of the three types of surface interactions in circular cylindrical geometry. The surface not only has an equilibrium curvature, but is also subject to accelerations associated with a steady rigid-body equilibrium rotation.

There are two reasons for wishing to study systems of this type. The great interest in stabilizing MH systems in connection with the thermonuclear program has already

been mentioned. The antiduality of MH and EH systems, which has been pointed out elsewhere,[5] should serve to show that study of the EH-If system can lead to a better understanding of the MH counterpart. An experimental investigation of the EH problem is possible, whereas the MH system can be considered experimentally only with great difficulty. In the second place, the study of cylindrical field coupled

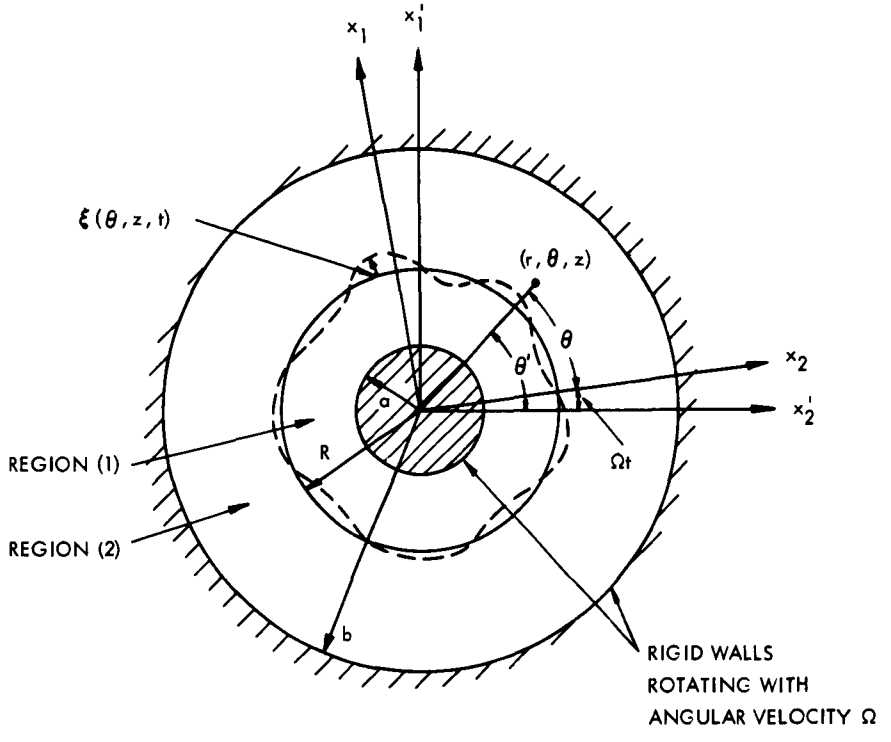


Fig. VIII-5. Cross section of rotating structure showing dimensions.

systems has always been of primary interest in spite of the fact that it is very difficult to achieve, experimentally, an equilibrium cylindrical geometry that approaches the idealization of a cylinder. For example, a jet in free fall changes in radius radically over a short distance unless the velocity is made high enough to produce turbulence. In fact, the very velocity of the jet is not a trivial addition to the dynamics, as has been shown in dealing with the convective systems.[6] In contrast, here we shall remove the effect of rotational convection by considering all interactions, both theoretically and experimentally, in the rotating frame.

2. Description of Problems

The mechanical configuration that will be considered here is shown in Fig. VIII-5. The region between two coaxial rigid circular cylinders is filled with two inviscid, incompressible nonmiscible fluids in such a way that they form an equilibrium interface at a radius R. The cross section is assumed to remain constant. The complete assembly is then assumed to rotate with an equilibrium rotational velocity Ω. It is possible to produce such a configuration by filling a cylinder with two fluids and rotating the cylinder fast enough to centrifugally drive the heavier fluid to the outside. The essentially rigid body rotation that results depends on the centrifugal acceleration being large compared with the acceleration of gravity. Hence, this last acceleration will be ignored in this analysis.

Although it is not usual to view them as such, problems of this kind are characterized by a discrete set of possible transverse modes so that they comprise a type of surface resonator.[7] This is most apparent if there is no axial dependence of the modes considered. In the experiments described in this report, the technique used to verify the theoretical considerations will take advantage of this fact. In any case, an experiment must involve a finite structure so that the axial modes will also be selected from a discrete set of wave numbers.

Here, both EH-I and EH-II interactions will be considered. The type EH-1 interaction is produced by an electric field that, in general, is in the radial direction. The EH-Ip is produced by equipotentials on the inner and outer cylinders because the region between would be filled with dielectric. However, in the EH-If interaction either the inner or outer cylinder can form one of the equipotentials, while the other might be formed by the interface. For the EH-II interaction, the inner and outer cylinders must be dielectrics, with the electric field imposed by parallel plates that are perpendicular to the z-axis. The types of velocity and electric fields to be considered are given by

$$\overline{V}' = V_1'\overline{a}_1 + (\Omega r + V_2')\,\overline{a}_2 + V_3'\overline{a}_3 \tag{1}$$

$$\left.\begin{array}{l} \text{EH-If} \\ \text{EH-Ip} \end{array}\right\} \overline{E} = \left(\frac{v}{r}+e_1\right)\overline{a}_1 + e_2\overline{a}_2 + e_3\overline{a}_3$$

$$\text{EH-II}\} \quad \overline{E} = e_1\overline{a}_1 + e_2\overline{a}_2 + (E_o+e_3)\,\overline{a}_3 \qquad \Bigg\} \tag{2}$$

3. E-H Equations of Motion

The EH bulk equations will be taken as

$$\nabla \cdot \overline{V}' = 0 \tag{3}$$

$$\rho \frac{D\overline{V}'}{Dt} + \nabla p' = 0 \tag{4}$$

$$\nabla \times \overline{E} = 0 \tag{5}$$

$$\nabla \cdot \epsilon \overline{E} = 0, \tag{6}$$

where p' is the total pressure given by

$$p' = P + \frac{1}{2}\epsilon E^2.$$

The primes denote quantities evaluated in the rest frame. The velocities of interest are not large enough to require any distinction between frames for the electric fields.

It is apparent that the mechanics would be considerably simplified by viewing the motion from the rotating frame. For this reason, Eqs. 3 and 4 will now be transformed to the rotating frame. This is accomplished by the transformation

$$\begin{aligned} r &= r' & V_1' &= V_1 \\ \vartheta &= \theta' - \Omega t & V_2' &= V_2 + \Omega r \\ z &= z' & p' &= p_o + p \\ t &= t' \end{aligned} \tag{7}$$

If only linear terms are retained, the momentum and continuity equations require

$$p_o = \frac{1}{2}\rho\Omega^2 r^2 + \text{constant} \tag{8}$$

$$\rho\left[\frac{\partial V_1}{\partial t} - 2\Omega V_2\right] + \frac{\partial p}{\partial r} = 0 \tag{9}$$

$$\rho\left[\frac{\partial V_2}{\partial t} + 2\Omega V_1\right] + \frac{1}{r}\frac{\partial p}{\partial \theta} = 0 \tag{10}$$

$$\rho\left[\frac{\partial V_3}{\partial t}\right] + \frac{\partial p}{\partial z} = 0 \tag{11}$$

$$\frac{1}{r}\frac{\partial r V_1}{\partial r} + \frac{1}{r}\frac{\partial V_2}{\partial \theta} + \frac{\partial V_3}{\partial z} = 0. \tag{12}$$

The terms in these equations are those that would be observed by a rotating observer. The usual techniques are now used to handle these equations. The total pressure is used as a potential for deriving the other mechanical variables, while the axial electric field is used to determine the remaining field components.

Because of the cylindrical geometry, the eigenvalues satisfying the axial boundary conditions can be chosen from those given by the assumed solution,

$$p = \hat{p}\, e^{j(\omega t + m\theta + kz)}. \tag{13}$$

Here, ω is the frequency as viewed in the moving frame. In all present considerations care will be taken to excite and measure all frequencies in the moving frame.

If the surface of the interface is to be continuous azimuthally, the eigenvalue m must have integer values. The wave number k will be determined in the usual way from the boundary conditions in the axial direction.

The assumed solution reduces the bulk equations of motion given by Eqs. 5, 6, and 9-12 to a pair of ordinary differential equations that define all of the remaining functions.

$$r^2 \frac{d^2\hat{p}}{dr^2} + r \frac{d\hat{p}}{dr} - \hat{p}[m^2 + r^2k^2\Delta^2] = 0, \tag{14}$$

where

$$\Delta^2 = \left[1 - \frac{4\Omega^2}{\omega^2}\right]$$

$$\frac{1}{r}\frac{dr}{dr}\left(\frac{d\hat{e}_3}{dr}\right) - \hat{e}_3\left[\frac{m^2}{r^2} + k^2\right] = 0. \tag{15}$$

Here, in terms of $\hat{p}$ and $\hat{e}_3$, the remaining functions are

$$\hat{V}_1 = \frac{j}{\rho\omega\Delta^2}\frac{d\hat{p}}{dr} + \frac{2jm\Omega}{\Delta^2\rho\omega^2}\left(\frac{\hat{p}}{r}\right) \tag{16}$$

$$\hat{V}_2 = \frac{-m}{\rho\omega\Delta^2}\left(\frac{\hat{p}}{r}\right) - \frac{2\Omega}{\Delta^2\rho\omega^2}\frac{d\hat{p}}{dr} \tag{17}$$

$$\hat{V}_3 = \frac{-k\hat{p}}{\rho\omega} \tag{18}$$

$$\hat{e}_2 = \frac{m}{k}\left(\frac{\hat{e}_3}{r}\right) \tag{19}$$

$$\hat{e}_1 = \frac{-j}{k}\frac{d\hat{e}_3}{dr}. \tag{20}$$

It is convenient to use as solutions to Eqs. 14 and 15 the Bessel function of first kind and the Hankel function of first kind.

$$\hat{p} = A_1 J_m(jk\Delta r) + A_2 H_m(jk\Delta r) \tag{21}$$

$$\hat{e}_3 = C_1 J_m(jkr) + C_2 H_m(jkr). \tag{22}$$

However, we note the important fact that the arguments of these functions as they occur in all of the mechanical variables depend on both the frequency ω and the

$$\nabla \times \overline{E} = 0 \tag{5}$$

$$\nabla \cdot \epsilon \overline{E} = 0, \tag{6}$$

where p' is the total pressure given by

$$p' = P + \frac{1}{2}\epsilon E^2.$$

The primes denote quantities evaluated in the rest frame. The velocities of interest are not large enough to require any distinction between frames for the electric fields.

It is apparent that the mechanics would be considerably simplified by viewing the motion from the rotating frame. For this reason, Eqs. 3 and 4 will now be transformed to the rotating frame. This is accomplished by the transformation

$$\begin{aligned} r &= r' & V_1' &= V_1 \\ \theta &= \theta' - \Omega t & V_2' &= V_2 + \Omega r \\ z &= z' & p' &= p_o + p \\ t &= t' \end{aligned} \tag{7}$$

If only linear terms are retained, the momentum and continuity equations require

$$p_o = \frac{1}{2}\rho\Omega^2 r^2 + \text{constant} \tag{8}$$

$$\rho\left[\frac{\partial V_1}{\partial t} - 2\Omega V_2\right] + \frac{\partial p}{\partial r} = 0 \tag{9}$$

$$\rho\left[\frac{\partial V_2}{\partial t} + 2\Omega V_1\right] + \frac{1}{r}\frac{\partial p}{\partial \theta} = 0 \tag{10}$$

$$\rho\left[\frac{\partial V_3}{\partial t}\right] + \frac{\partial p}{\partial z} = 0 \tag{11}$$

$$\frac{1}{r}\frac{\partial r V_1}{\partial r} + \frac{1}{r}\frac{\partial V_2}{\partial \theta} + \frac{\partial V_3}{\partial z} = 0. \tag{12}$$

The terms in these equations are those that would be observed by a rotating observer. The usual techniques are now used to handle these equations. The total pressure is used as a potential for deriving the other mechanical variables, while the axial electric field is used to determine the remaining field components.

Because of the cylindrical geometry, the eigenvalues satisfying the axial boundary conditions can be chosen from those given by the assumed solution,

$$p = \hat{p}\, e^{j(\omega t + m\theta + kz)}. \tag{13}$$

wave number k. This is the salient complication of the rotation.

The surface dynamics are represented in terms of the unit normal, n, directed in the positive r direction in equilibrium, and the surface displacement $\xi(\theta, z, t)$ from an equilibrium position R.

$$\frac{D}{Dt}\left[r-(\xi(\theta, z, t)+R)\right] = 0. \tag{23}$$

Remember that ξ is the surface perturbation as viewed in the fixed frame. Hence, Eq. 21 must be transformed to the rotating frame in which it takes on the expected nonconvective form

$$\hat{\xi} = -\frac{j\hat{V}_1}{\omega}. \tag{24}$$

The normal vector does not depend on rates of change with respect to time, and is therefore unaffected by the transformation to the rotating frame of reference. Hence

$$\left.\begin{aligned}\hat{n}_2 &= \frac{-m}{R\omega}\hat{V}_1\\ \hat{n}_3 &= \frac{-k}{\omega}\hat{V}_1\end{aligned}\right\}. \tag{25}$$

The boundary conditions are given here for convenience. They are of the same form as those used in conjunction with variables defined in the inertial frame because the transformation of all of the variables involved in the boundary conditions (see Eq. 7) is one to one. Hence, the definition of the interface and the conservation of momentum at the interface require that

$$\bar{n} \cdot \left[\bar{V}^{(2)}-\bar{V}^{(1)}\right] = 0 \tag{26}$$

$$n_a T\left(\frac{1}{R} - \frac{1}{R^2}\left(\xi + \frac{\partial^2\xi}{\partial\theta^2}\right) - \frac{\partial^2\xi}{\partial z^2}\right) - n_\beta\left[T^{(2)}_{a\beta}-T^{(1)}_{a\beta}\right] = 0, \tag{27}$$

where T is the surface tension, and $T_{a\beta}$ is the total stress tensor given by

$$T_{a\beta} = \epsilon E_a E_\beta - \delta_{a\beta}\left[\frac{\epsilon E_\gamma E_\gamma}{2} + p\right].$$

Also, the tangential component of the electric field must be continuous at the interface.

$$n \times \left[\bar{E}^{(2)}-\bar{E}^{(1)}\right] = 0. \tag{28}$$

If there is no free charge on the interface, a type I interaction requires the use of the additional condition

$$\bar{n} \cdot \left[\epsilon^{(2)}\bar{E}^{(2)}-\epsilon^{(1)}\bar{E}^{(1)}\right] = 0. \tag{29}$$

At r = a, b, the normal velocity must vanish because of the condition imposed by a rigid boundary. In the EH-II interaction it is assumed that these boundaries are composed of a dielectric material so that it is possible to support an axial electric field that is uniform in the unperturbed state. The walls are assumed to have the same dielectric properties as the adjacent fluid dielectric, so that there is no polarization charge or Korteweg current on the rigid boundaries. On the other hand, the type I interactions are assumed to result from equipotentials at r = a, b; and this requires that $e_3 = 0$ there.

4. Dispersion Equations

The technique of finding the dispersion equations that are consistent with perturbations of the form of Eq. 13 is now straightforward. Equations 16-20, 24, and 25 make it possible to evaluate all of the dependent variables in terms of the solutions given by Eqs. 21 and 22, that is, in terms of eight constants, $A_1^{(1)}$ $A_2^{(1)}$, $C_1^{(1)}$ $C_2^{(1)}$, $A_1^{(2)}$ $A_2^{(2)}$, $C_1^{(2)}$ $C_2^{(2)}$. (The superscripts denote the region in which the variable is to be evaluated.)

The dispersion relation for each of the problems is the compatibility condition of these eight constants in the equations resulting when the solutions are substituted in the boundary conditions of Eqs. 26-29.

The cylindrical geometry poses the complication of a radial dependence for the equilibrium pressure field; and in the type I interactions, a radial dependence for the equilibrium electric fields. Hence, in this linear analysis, these respective fields must be written in the form

$$P = \text{constant} + \frac{1}{2}\rho\Omega^2(R+\xi)^2 + p(r,\theta,z,t)$$

$$\approx \text{constant} + \frac{1}{2}\rho\Omega^2(R^2+2R\xi) + p(r,\theta,z,t) \tag{30}$$

$$E_1 = \frac{V}{(R+\xi)} + e_1 = \frac{V}{R} - \frac{V}{R^2}\xi + e_1. \tag{31}$$

Investigation shows that for the EH-If interaction, Eqs. 26-29 are linear combinations of Eqs. 26, $(27)_1$, $(27)_3$, $(28)_2$ and the condition that $v_1 = 0$ at r = a, b. One of these equations ceases to be independent in the limit where there is no free charge on the interface. In this case (the EH-Ip interaction) Eq. $(27)_3$ is replaced by Eq. 29.

The EH-II interaction is defined by Eqs. 26, $(27)_1$, $(27)_3$, 28, and the condition that $v_1 = 0$ at r = a, b. It is assumed that all electromagnetic perturbation fields are finite at infinity and at the origin. This last condition replaces two equations, in that it requires $C_1^{(2)} = C_2^{(1)} = 0$. The dispersion equations for each of the three problems can be summarized as follows:

$$\omega^2 = \frac{T}{R^2\rho_{eq}}\{(kR)^2-(1-m^2)+F+G\}, \tag{32}$$

where

$$F = \frac{R^3\Omega^2}{T}\left(\rho^{(2)}-\rho^{(1)}\right)$$

$$\rho_{eq} = \Delta\left[\rho^{(1)}P_m(jk\Delta a, jk\Delta R)-\rho^{(2)}P_m(jk\Delta b, jk\Delta R)\right]$$

$$P_m(x, y) = \frac{1}{jkR}\left[\frac{J_m(y)\,L_m(x) - Q_m(x)\,H_m(y)}{L_m(x)\,Q_m(y) - Q_m(x)\,L_m(y)}\right]$$

$$Q_m(x) = J'_m(x) + \frac{2m\Omega}{\omega}\,\frac{J_m(x)}{x}$$

$$L_m(x) = H'_m(x) + \frac{2m\Omega}{\omega}\,\frac{H_m(x)}{x}$$

$$S_m(kR, kb) = jkR\left\{\frac{H'_m(jkR)\,J_m(jkb) - H_m(jkb)\,J'_m(jkR)}{H_m(jkR)\,J_m(jkb) - J_m(jkR)\,H_m(jkb)}\right\}$$

$$T_m(kR, ka) = jkR\left[\frac{J'_m(jkR)\,H_m(jka) - J_m(jka)\,H'_m(jkR)}{J_m(jkR)\,H_m(jka) - J_m(jka)\,H_m(jkR)}\right]$$

$$\Delta = \sqrt{1 - \frac{4\Omega^2}{\omega^2}}$$

for all of the problems, but G depends on the type of interaction that is considered.

EH-If

$$G = \Gamma^{(2)}[S_m(kR, kb)+1] - \Gamma^{(1)}[T_m(kR, ka)+1]$$

$$\Gamma = \frac{v^2\epsilon}{RT}.$$

EH-Ip

$$G = \frac{1}{RT}\left\{\frac{v^{(1)}\,v^{(2)}\left(\epsilon^{(2)}-\epsilon^{(1)}\right)^2}{\dfrac{\epsilon^{(1)}}{S_m(kR, kb)} - \dfrac{\epsilon^{(2)}}{T_m(kR, ka)}} + \epsilon^{(2)}\left(v^{(2)}\right)^2 - \epsilon^{(1)}\left(v^{(1)}\right)^2\right\}.$$

EH-II

$$G = \frac{\left(\epsilon^{(2)}-\epsilon^{(1)}\right)(Rk)^2 E_o^2 R}{T\left[\epsilon^{(1)}(jkR)\frac{J'_m(jkR)}{J_m(jkR)} - \epsilon^{(2)}(jkR)\frac{H'_m(jkR)}{H_m(jkR)}\right]}.$$

If modes of motion are considered that have no z dependence, then the dispersion equations take a simple and convenient form that is quadratic in the frequency ω. To show this, it is recognized that as $k \to 0$ (or as perturbations in the axial direction assume an infinite wave length),

$$T_m(kR, ka) \to m\left[\frac{\left(\frac{R}{a}\right)^{2m}+1}{\left(\frac{R}{a}\right)^{2m}-1}\right] \tag{33}$$

$$S_m(kR, kb) \to m\left[\frac{\left(\frac{b}{R}\right)^{2m}+1}{\left(\frac{b}{R}\right)^{2m}-1}\right]$$

$$P_m(x, y) \to \frac{1}{m\Delta}\left[\frac{\left(1+\frac{2\Omega}{\omega}\right)+\left(\frac{y}{x}\right)^{2m}\left(1-\frac{2\Omega}{\omega}\right)}{\left(\frac{y}{x}\right)^{2m}-1}\right]$$

and the dispersion equations become

$$A\left(\frac{\omega}{\Omega}\right)^2 + 2B\left(\frac{\omega}{\Omega}\right) + C + (1-m^2) - F = 0, \tag{34}$$

where

$$A = \frac{F}{m\left(\rho^{(2)}-\rho^{(1)}\right)}\left\{\left[\frac{\left(\frac{R}{a}\right)^{2m}+1}{\left(\frac{R}{a}\right)^{2m}-1}\right]\rho^{(1)} - \left[\frac{\left(\frac{R}{b}\right)^{2m}+1}{\left(\frac{R}{b}\right)^{2m}-1}\right]\rho^{(2)}\right\}$$

$$B = \left(\frac{F}{m}\right).$$

EH-If

$$C = \left(\Gamma^{(1)}-\Gamma^{(2)}\right) + m\left[\Gamma^{(2)}\left[\frac{\left(\frac{b}{R}\right)^{2m}+1}{\left(\frac{b}{R}\right)^{2m}-1}\right] + \Gamma^{(1)}\left[\frac{\left(\frac{R}{a}\right)^{2m}+1}{\left(\frac{R}{a}\right)^{2m}-1}\right]\right].$$

EH-Ip

$$C = \left(\Gamma^{(1)}-\Gamma^{(2)}\right) - \frac{m^2 v^{(2)} v^{(1)}\left(\epsilon^{(1)}-\epsilon^{(2)}\right)^2}{RT\left[\epsilon^{(1)}\frac{\left(1-\left(\frac{R}{b}\right)^{2m}\right)}{\left(1+\left(\frac{R}{b}\right)^{2m}\right)} + \epsilon^{(2)}\frac{\left(1-\left(\frac{a}{R}\right)^{2m}\right)}{\left(1+\left(\frac{a}{R}\right)^{2m}\right)}\right]}.$$

EH-II

$$C = 0.$$

Note that for a given mode (a given integer value of m) A, B, and C are constants fixed by the physical parameters and the geometry. For each mode, there are two possible frequencies of oscillation. In the limit $\Omega \to 0$, these two values of ω approach the same value. Hence, the effect of the rotation is to split the possible resonant frequencies into disparate pairs. The EH-II waves couple to the electric field only if the direction of propagation is in the direction of the equilibrium electric field.[2] It is therefore not surprising that the electric field has no influence on the resonant frequency of the type II resonator if there is no axial dependence of the surface deformations.

5. Experiment

The EH-If waves discussed in the previous sections are being investigated by using the apparatus shown in Fig. VIII-6. The detailed drawing shows the resonator that is rotated with the frequency Ω. This rotation is provided by a dc motor equipped with a feedback loop for speed control. (The measurement of the resonant frequency ω is sensitive enough to require a speed controlled to four significant figures.) The speed is measured within 1/8 of a revolution with a 10-sec sample time, as shown in Fig. VIII-6. An excitation of the surface waves was provided by torque pulsations from a dc motor with its field excited by a square wave having a repetition rate $\omega/2\pi$. Because the oscillator that generated the square wave operates at 16 times the frequency ω, it is possible to measure the excitation frequency to four places, by using a 10-sec sample time.

The low-frequency m = 1 and m = 2 modes have been measured by using a cylinder of length 8.75 in. with a = 0.5 in., b = 1.867 in. and R = 1.31 in. Although frequency shifts with electric field have been observed (they are small), attention has been confined thus far to attempting to understand the waves without field coupling. Figure VIII-7

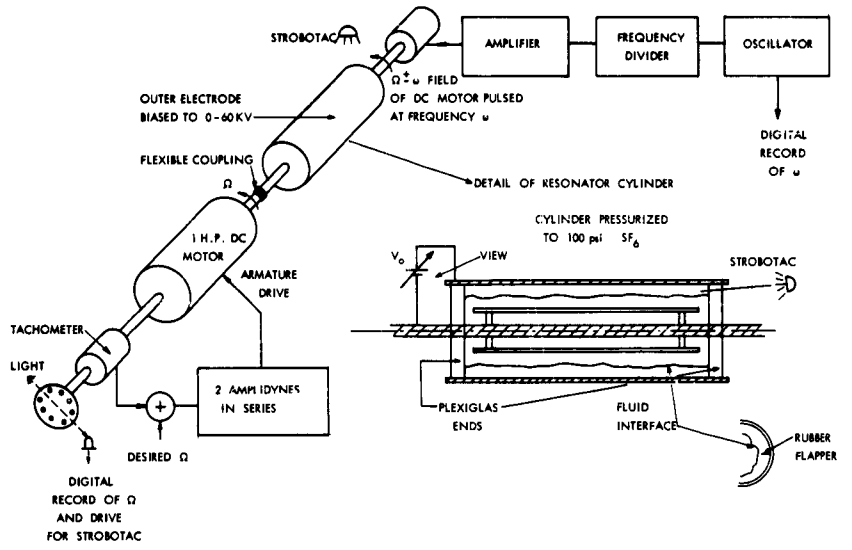


Fig. VIII-6. Experimental arrangement for study of EH surface waves on rotating interface.

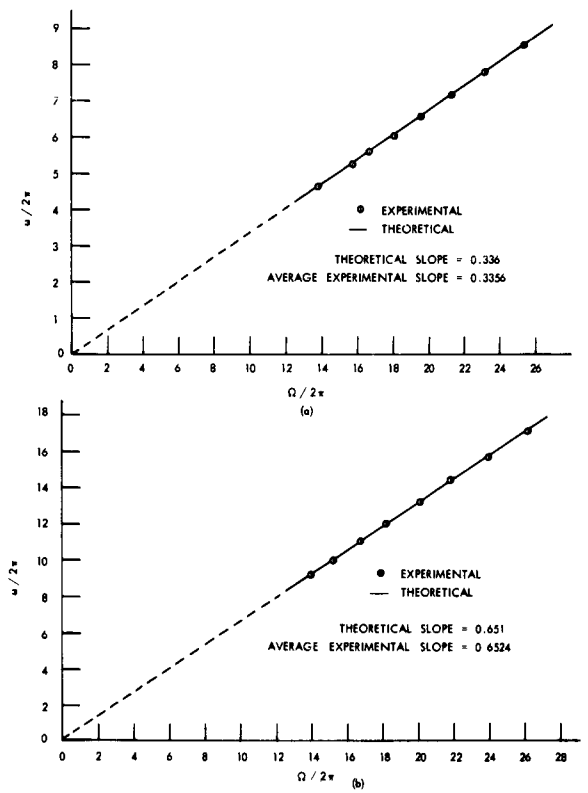


Fig. VIII-7.

(a) Low-frequency m = 1, k = 0 mode (no electric field).
(b) Low-frequency m = 2, k = 0 mode (no electric field).

shows typical data with no electric field. It can be seen that the experiment justifies the mechanical model used (at least for the prediction of the resonant frequency). The solid lines were drawn by using Eq. 34. The surface waves are excited on the interface of the fluid by a thin rubber "flapper" that is attached to the outer wall as shown in Fig. VIII-6. The rotational pulsations of the cylinder then set the fluid into a slight motion which is observed under stroboscopic light. An extremely sensitive indication of resonance is given when the resonant motions of the interface are accompanied by no deflection of the flapper. It is at this frequency that the flapper does not inhibit the motion of the fluid, and hence "pull" the resonant frequency.

J. R. Melcher

References

1. S. Chandrasekhar, Hydrodynamic and Hydromagnetic Stability (Oxford University Press, London, 1961), p. 428.

2. J. R. Melcher, Electrohydrodynamic and magnetohydrodynamic surface waves and instabilities, Phys. Fluids 4, 11 (1961).

3. H. Lamb, Hydrodynamics (Dover Publications, Inc., New York, 1932), p. 307.

4. S. Chandrasekhar, op. cit., p. 284.

5. J. R. Melcher, Field-Coupled Surface Waves (unpublished).

6. J. R. Melcher, Electrohydrodynamic surface resonators, Phys. Fluids 5, 9 (1962).

D. MAGNETOACOUSTIC-WAVE EXPERIMENT

1. Introduction

It has been shown by Haus[1] that it is theoretically possible to use magnetoacoustic waves to convert kinetic power of a flowing plasma to ac electrical power by means of an external, magnetically coupled circuit. In order for this amplifying interaction to occur, the following conditions must be satisfied: (a) the plasma must be flowing through a region of constant applied transverse magnetic field; (b) the flow velocity must be greater than the magnetoacoustic wave velocity, and (c) the plasma electrical conductivity must be sufficiently large to permit the waves to amplify as they propagate.

An experiment has been devised to try to produce a plasma in a system satisfying these conditions. The system is a homopolar device (see Fig. VIII-8) in which a gas is heated and accelerated electrically in a uniform externally applied magnetic field. We are trying to determine whether or not the resulting rotating plasma will satisfy the above-given conditions.

In recent years there has been a growing interest in rotating plasmas for such uses as fast discharge capacitors,[2] arc gas heaters for wind-tunnel or materials studies,[3]

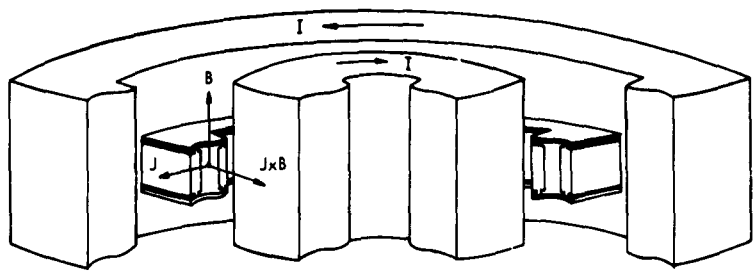


Fig. VIII-8. Homopolar device.

and for basic studies in plasma physics.[4] In each of the cases reported the magnetic field was so large that the flow velocity was less than the magnetoacoustic-wave propagation velocity. Furthermore, there were no attempts to study magnetoacoustic waves. Nevertheless these studies have provided a useful starting point and some theoretical framework for the experiments reported on here. Some initial experiments will be described in which the objective was to produce a plasma satisfying the conditions discussed above.

2. Description of Experiment

The apparatus for the experiment is shown in Fig. VIII-9. The channel (Fig. VIII-10) consists of two concentric circular brass rings. The top and bottom cover plates are either glass or Plexiglas. The inside surfaces of the brass electrodes are notched near the insulating walls and coated with sauereisen, a high-temperature porcelainlike material, in an attempt to confine the plasma away from the insulating walls. The electrode currents are carried by 4 radial arms that join at the axis beneath the electrodes. The chamber is sealed by 4 neoprene O-rings. No external support is required for the cover plates once the chamber is evacuated. The axial magnetic field is produced by two concentric coils connected in series; the field in the region between the coils is uniform within ±5 per cent. Five taps are provided on each coil to vary the amount of inductance from 3-30 mh. The energy for powering the field coils and the plasma is supplied by an 800-μf, 4-kv capacitor bank. The duration of an experiment is 10 msec, or less, depending on the parameters of the system, so that no external cooling is necessary for either the coils or channel. The switching elements used in the experiments are type 5550 ignitrons, which are air-cooled and rated at 12 kv and 20 coulombs for pulse work,[5] or, for example, 20,000 amps for 1 msec. The ignitron igniters are excited by 5C22 hydrogen thyratrons that are fired by delay circuits and pulse-forming networks. See Fig. VIII-11.

The channel is evacuated to approximately 1 μ Hg pressure before each firing and

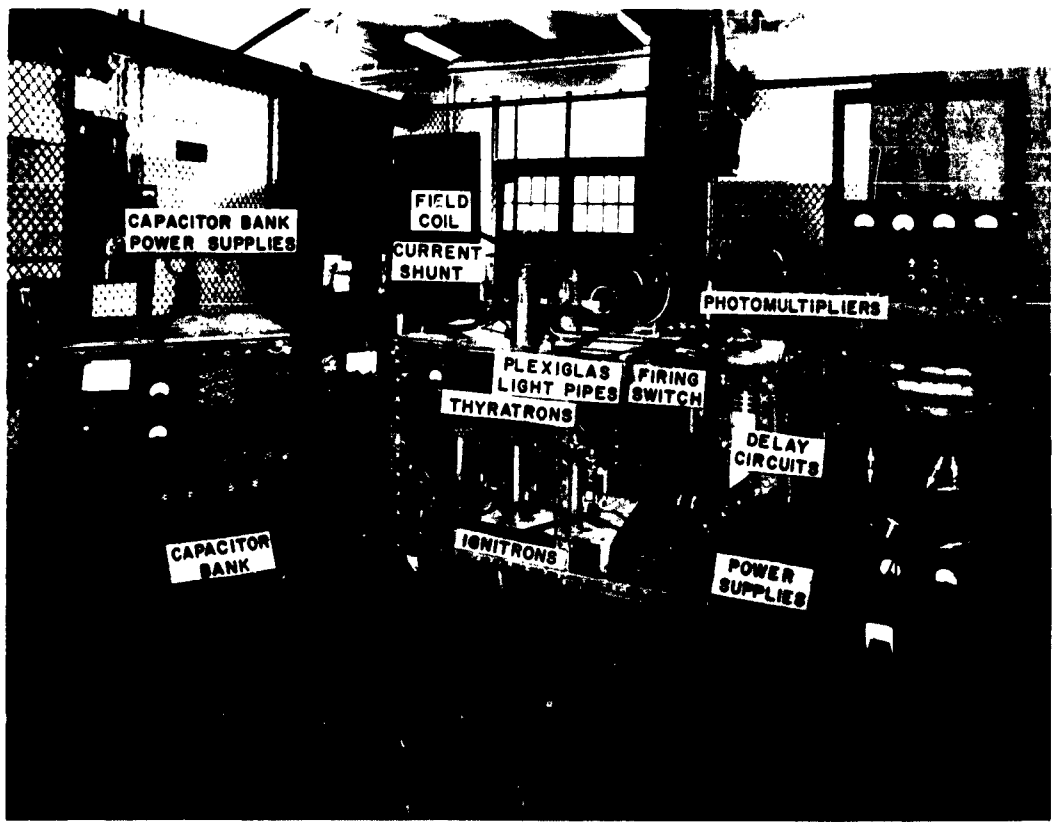


Fig. VIII-9. Rotating-plasma apparatus.

backfilled to the desired pressure with the working fluid. Most of the experiments were performed with helium in the pressure range 0.5-10 mm Hg. At pressures much out of this range it was not possible to break down the gas with the available capacitor bank when an external magnetic field was applied.

Several channel configurations had been used in previous experiments, which formed the basis for the present design. Two circuit configurations have been studied: (a) a series connection in which the field coils are in series with the plasma and (b) separately excited circuits in which the channel and field coils are pulsed from separate capacitor banks. The series arrangement has the advantages that it (a) uses the coil inductance for current limiting, (b) provides a reasonably long discharge time and (c) produces high magnetic fields. A peak discharge current of 1400 amps and magnetic field of 15,000 gauss is possible with a discharge time of 5 msec. At lower field and discharge currents an arc is observed to accelerate azimuthally until a point is reached at which

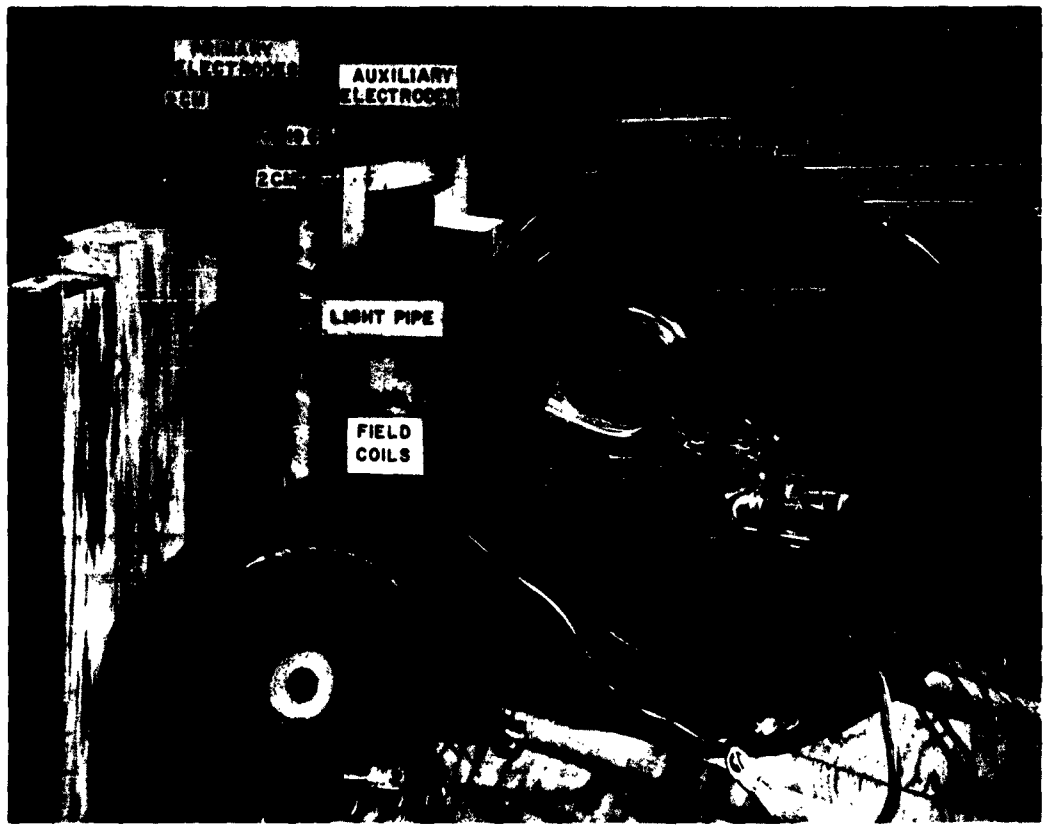


Fig. VIII-10. Channel with field coils removed.

time the arc suddenly switches mode and becomes diffuse for most of the remainder of the discharge. The current, voltage, and light intensity are illustrated in Fig. VIII-12. The only noted effect of pressure changes in the range 0.5-10 mm Hg was that the plasma was not diffuse as long during the discharge time at the higher pressures. From the

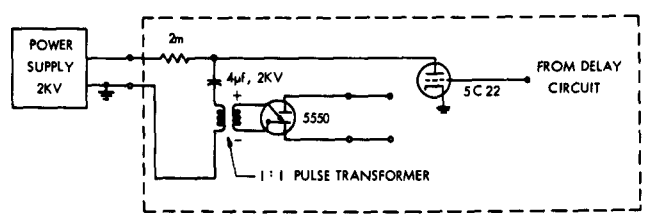


Fig. VIII-11. Typical ignitron firing circuit.

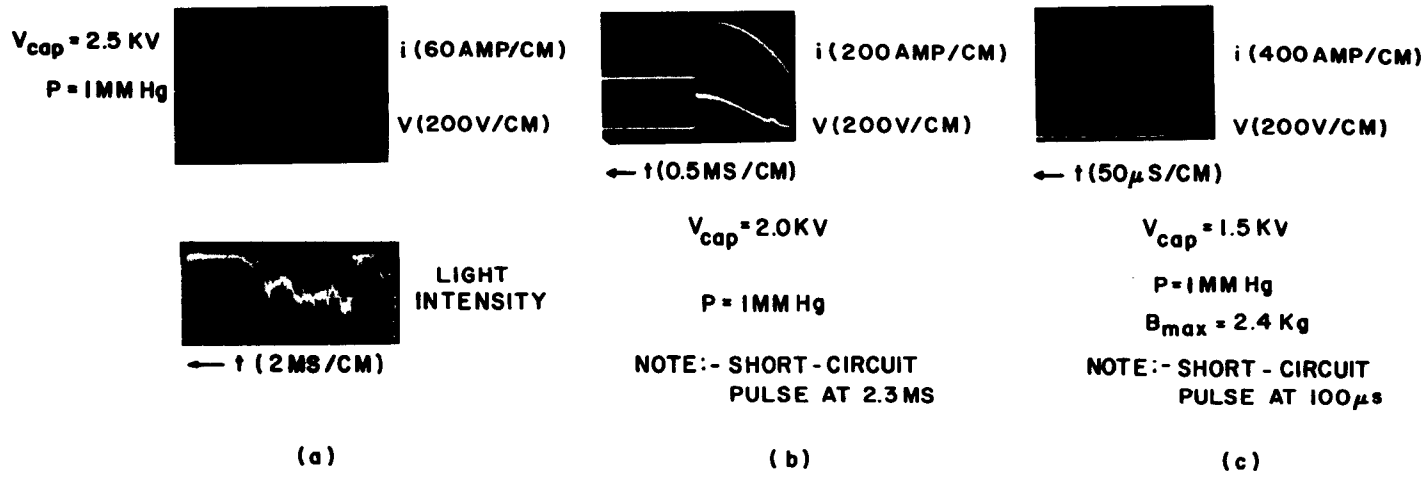


Fig. VIII-12. Waveforms of rotating plasma.

voltage and current traces the voltage is quite nearly proportional to the current. This would be true, for example, if the voltage was produced by a $\vec{v} \times \vec{B}$ back emf if we assume a limiting velocity for the plasma. The velocity corresponding to the values given is of the order of the ionization energy velocity, a result shown by others.[6, 7] Attempts to short-circuit the plasma to measure the charge stored in the plasma capacitor, however, did not show any measurable stored charge.

If the inductance of the field coils is reduced, then with a peak field of 6500 gauss and a discharge current of 600 amps, short-circuiting the plasma resulted in the recovery of approximately 100 μcoulomb of charge. This is equivalent to an ionization of 0.1 per cent (or $n_i = 10^{13}$ ions/cc) if it is assumed that the rotational energy of the ions is of the order of the ionization energy. Observations of the light intensity indicate that at the higher currents and magnetic fields the rotating arc phase of the discharge no longer appears.

With the separately excited configuration used and with a current-limiting resistor of 0.6 ohm in the plasma circuit, experiments indicate that the plasma again exhibits capacitive properties (Fig. VIII-12c). In these experiments the magnetic field is first established by part of the capacitor bank and the remainder of the bank is discharged into the channel when the field is maximum.

At the present time we are proceeding with instrumentation to determine conditions experimentally in the rotating plasma. Attempts to launch magnetoacoustic waves in the rotating plasma have been inconclusive.

F. D. Ketterer

References

1. H. A. Haus, J. Appl. Phys. 33, 2161-2172 (1962).

2. O. Anderson, W. Baker, A. Bratenahl, H. Furth, and W. Kunkel, J. Appl. Phys. 30, 188 (1959).

3. R. Mayo and D. Davis, Jr., American Rocket Society Electric Propulsion Conference, Berkeley, California, March 1962.

4. K. Halbach, W. Baker, and R. Layman, Phys. Fluids 5, 1482 (1962).

5. D. B. Hopkins and P. F. Pellissier, UCRL-9968, Lawerence Radiation Laboratory, University of California, Berkeley, Calif., December 11, 1961.

6. H. Alfvén, Res. Modern Phys. 32, 710 (1960).

7. U. V. Fahleson, Phys. Fluids 4, 123 (1961).

COMMUNICATION SCIENCES AND ENGINEERING

IX. STATISTICAL COMMUNICATION THEORY*

Prof. Y. W. Lee	M. E. Austin	P. L. Konop
Prof. A. G. Bose	R. F. Bauer	A. J. Kramer
Prof. D. J. Sakrison	E. M. Bregstone	D. E. Nelsen
Prof. M. Schetzen	J. D. Bruce	J. K. Omura
Prof. H. L. Van Trees, Jr.	A. M. Bush	A. V. Oppenheim
V. R. Algazi	J. K. Clemens	R. B. Parente
R. Alter	A. G. Gann	W. S. Smith
D. S. Arnstein	C. E. Gray	D. W. Steele
	T. G. Kincaid	

A. GENERALIZATION OF THE ERROR CRITERION IN NONLINEAR THEORY BASED ON THE USE OF GATE FUNCTIONS

In the Wiener theory of optimum nonlinear systems, the measure of performance is the mean-square error and the input is a Gaussian process. A reason for the choice of this error criterion and type of input is the resulting relative analytical and experimental simplicity by which a nonlinear system can be determined. Recently, experimental procedures have been studied by which one can determine optimum nonlinear systems for error criteria other than the mean-square error criterion and inputs other than a Gaussian process.[1,2] The basic procedure is to expand the class of systems of interest into a complete set of operators, $\mathcal{K}_n$, so that the output of any system of this class can be expressed as

$$y(t) = \sum_{n=1}^{N} a_n y_n(t) \tag{1}$$

in which

$$y_n(t) = \mathcal{K}_n[x(t)]. \tag{2}$$

Figure IX-1 is a schematic representation of Eq. 1. In this representation, the coefficients, a_n, are amplifier gains. The determination of an optimum system is thus reduced to the determination of these coefficients. The procedure is then to measure experimentally the desired function of the error and simply to adjust the amplifier gains in order to minimize this quantity. This procedure is guaranteed to result in the optimum system of the class being represented for convex error criteria, since we are then guaranteed that there is a unique minimum of the function of the error and that there are no local minima. A difficulty with this method is that, in general, the amplifier

*This work was supported in part by the National Institutes of Health (Grant MH-04737-02); and in part by the National Science Foundation (Grant G-16526).

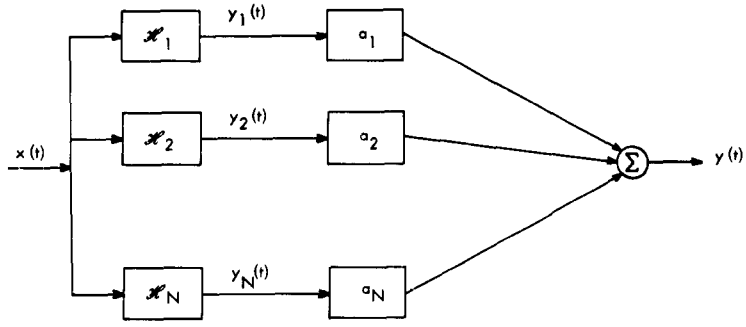


Fig. IX-1. Operator expansion of a system.

gain settings are not independent of one another and therefore an iterative adjustment procedure is required. This interdependence of the amplifier gain settings also makes any estimation of the optimum gain settings by analytical techniques too unwieldy to be obtained in any practical manner. Also, one generally has difficulty in obtaining a reasonable estimate of the class of nonlinear systems to represent. However, there is a set of operators based on the gate functions which can be used to expand the desired system with which, for any error criterion, the amplifier gain settings do not interact. In this report, we shall present some experimental and analytical techniques by which the amplifier gains can be determined when this set of operators is used. To explain the techniques, the determination of optimum nonlinear no-memory systems for various error criteria will be presented. The extension to nonlinear systems with memory will then be given.

1. The Gate Functions

For nonlinear no-memory systems, the set of operators that we shall use is the gate functions, Q_n. These functions were first introduced into the study of nonlinear systems by Bose.[3] For this set of operators and input, x(t), the outputs, $y_n(t)$, are

$$\left.\begin{aligned} y_0(t) = Q_0[x(t)] &= \begin{cases} 1 & \text{if } -\infty < x(t) < x_1 \\ 0 & \text{otherwise} \end{cases} \\ y_n(t) = Q_n[x(t)] &= \begin{cases} 1 & \text{if } x_n \leq x(t) < x_{n+1};\ n \neq 0, N \\ 0 & \text{otherwise} \end{cases} \\ y_N(t) = Q_N[x(t)] &= \begin{cases} 1 & \text{if } x_N \leq x(t) < \infty \\ 0 & \text{otherwise} \end{cases} \end{aligned}\right\} \tag{3}$$

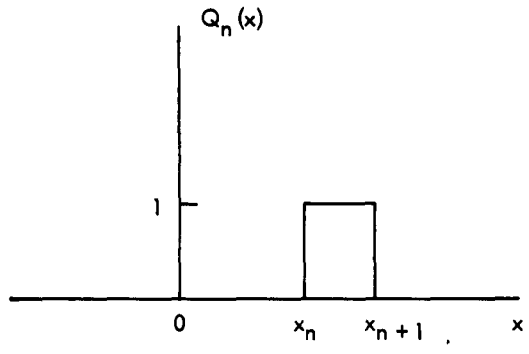


Fig. IX-2. Transfer characteristic of a gate function.

in which $x_m > x_n$ if $m > n$. Figure IX-2 is a plot of the n^{th} gate function. Any step approximation of a nonlinear no-memory system can be expressed in terms of this set as

$$y(t) = \sum_{n=0}^{N} a_n Q_n[x(t)], \tag{4}$$

in which we have not placed any restrictions on x(t).

Before proceeding, we note four important properties of the gate functions. The first is

$$\sum_{n=0}^{N} Q_n[x(t)] = 1. \tag{5}$$

The second is

$$Q_n^P[x(t)] = Q_n[x(t)] \qquad \text{for } P > 0. \tag{6}$$

The third is

$$\prod_{n=1}^{P} Q_{\alpha_n}[x(t)] = 0 \qquad \text{unless } \alpha_1 = \alpha_2 = \ldots = \alpha_P. \tag{7}$$

Then, by use of Eqs. 6 and 7, we obtain the fourth property, which is

$$\left[\sum_{n=0}^{N} a_n Q_n[x(t)]\right]^P = \sum_{n=0}^{N} a_n^P Q_n[x(t)]. \tag{8}$$

2. Optimum No-Memory Systems for Minimum $\overline{|E(t)|^P}$

In order to illustrate some of the properties of the gate functions and also methods for determining the coefficients, a_n, we first shall discuss the determination of optimum nonlinear no-memory systems for error criteria that are the mean P^{th} power of the magnitude of the error, E(t). That is, the error criteria that we first shall discuss are $\|E(t)\| = \overline{|E(t)|^P}$ for P = 1, 2, 3, ... We shall then generalize these results to error criteria that are arbitrary functions of the error.

For the input x(t), let the desired output be z(t). The error is then

$$E(t) = z(t) - \sum_{n=0}^{N} a_n Q_n[x(t)]. \tag{9}$$

For convenience, we shall omit writing the argument t. Thus we shall write the error as given by Eq. 9

$$E = z - \sum_{n=0}^{N} a_n Q_n[x] \tag{10}$$

with the understanding that E, x, and z are functions of time. Then the P^{th} power of the error is

$$E^P = \left[z - \sum_{n=0}^{N} a_n Q_n(x) \right]^P . \tag{11}$$

We shall obtain a more convenient form of this expression. Equation 11 can be expanded as

$$\begin{aligned} E^P = z^P &+ c_1 z^{P-1} \left[\sum_{n=0}^{N} a_n Q_n(x) \right] \\ &+ c_2 z^{P-2} \left[\sum_{n=0}^{N} a_n Q_n(x) \right]^2 + \cdots \\ &+ c_P \left[\sum_{n=0}^{N} a_n Q_n(x) \right]^P \end{aligned} \tag{12}$$

in which the coefficients, c_n, are the binomial coefficients. Substituting Eq. 8 in Eq. 12, we then have

$$E^P = z^P + c_1 z^{P-1} \sum_{n=0}^{N} a_n Q_n(x)$$

$$+ c_2 z^{P-2} \sum_{n=0}^{N} a_n^2 Q_n(x) + \ldots$$

$$+ c_P \sum_{n=0}^{N} a_n^P Q_n(x). \tag{13}$$

Equation 13 can be written in the form

$$E^P = z^P + \sum_{n=0}^{N} \left[c_1 z^{P-1} a_n + c_2 z^{P-2} a_n^2 + \ldots + c_P a_n^P \right] Q_n(x)$$

$$= z^P + \sum_{n=0}^{N} \left[(z-a_n)^P - z^P \right]$$

$$= z^P \left[1 - \sum_{n=0}^{N} Q_n(x) \right] + \sum_{n=0}^{N} [z-a_n]^P Q_n(x). \tag{14}$$

By use of Eq. 5, we then have

$$E^P = \sum_{n=0}^{N} [z-a_n]^P Q_n(x). \tag{15}$$

The magnitude of the P^{th} power of the error can be expressed in the form

$$|E|^P = [E^{2P}]^{1/2}. \tag{16}$$

Thus, from Eq. 15 we have

$$\overline{|E|^P} = \left[\overline{\sum_{n=0}^{N} (z-a_n)^{2P} Q_n(x)} \right]^{1/2}, \tag{17}$$

in which the bar indicates the time average of the function.

The optimum set of coefficients, a_n, for which Eq. 17 is a minimum can be determined by setting the derivative with respect to a_j equal to zero. Thus, from Eq. 17 we have

$$\frac{\partial \overline{|E|^P}}{\partial a_j} = -P\left[\frac{\overline{(z-a_j)^{2P-1}Q_j(x)}}{|E|^P}\right]. \tag{18}$$

The numerator in Eq. 18 is zero except when the amplitude of x(t) falls in the interval of the j^{th} gate function. At such times, the denominator is equal to $\left[(z-a_j)^{2P}Q_j(x)\right]^{1/2}$. Thus we can rewrite Eq. 18 to obtain

$$\frac{\partial \overline{|E|^P}}{\partial a_j} = -P\left[\frac{\overline{(z-a_j)^{2P-1}}}{|z-a_j|^P}Q_j(x)\right]$$

$$= -P\left[\frac{\overline{|z-a_j|^P}}{(z-a_j)}Q_j(x)\right]. \tag{19}$$

However,

$$(z-a_j) = |z-a_j|\,\mathrm{Sgn}(z-a_j) \tag{20}$$

in which

$$\mathrm{Sgn}(z-a_j) = \begin{cases} 1 & \text{if } z > a_j \\ 0 & \text{if } z = a_j \\ -1 & \text{if } z < a_j. \end{cases} \tag{21}$$

Substituting Eq. 20 in Eq. 19, we find that the condition for $\overline{|E|^P}$ to be a minimum is

$$\overline{|z-a_j|^{P-1}\mathrm{Sgn}(z-a_j)\,Q_j(x)} = 0. \tag{22}$$

We can show that Eq. 22 is the condition for a minimum by differentiating Eq. 18 with respect to a_j and substituting Eq. 22 in the result. Then we obtain

$$\frac{\partial^2 \overline{|E|^P}}{\partial a_j^2} = (2P-1)\,P\left[\frac{\overline{(z-a_j)^{2(P-1)}}}{|E|^P}Q_j(x)\right] \geq 0. \tag{23}$$

Equation 23 is positive, since $P \geq 1$ and all terms being averaged are always positive. Thus Eq. 22 is the condition for a minimum. We note that, by using the gate functions, the amplifier gain settings do not interact and they can be determined individually by means of Eq. 22.

Equation 22 provides a convenient experimental method for detcrmining the desired

coefficients, a_n. We note that if P is even, then Eq. 22 can be written in the form

$$\overline{(z-a_j)^{P-1}Q_j(x)} = 0 \qquad [P \text{ even}]. \tag{24}$$

A circuit for experimentally determining the value of a_j that satisfies Eq. 24 is depicted in Fig. IX-3. The battery voltage, V, is adjusted to make the meter read zero; the voltage, V, is then numerically equal to the optimum value of a_j. Another procedure is to expand Eq. 24 as

$$\overline{(z-a_j)^{P-1}Q_j(x)} = \sum_{n=0}^{P-1} c_n \overline{z^n Q_j(x)}\, a_j^{(P-1-n)} = 0, \tag{25}$$

in which the coefficients, c_n, are the binomial coefficients. The averages, $\overline{z^n Q_j(x)}$, can be measured experimentally. The optimum value of a_j can then be determined by substituting the measured values of $\overline{z^n Q_j(x)}$ in Eq. 25 and solving for the root of the resulting (P-1)-degree polynomial in a_j.

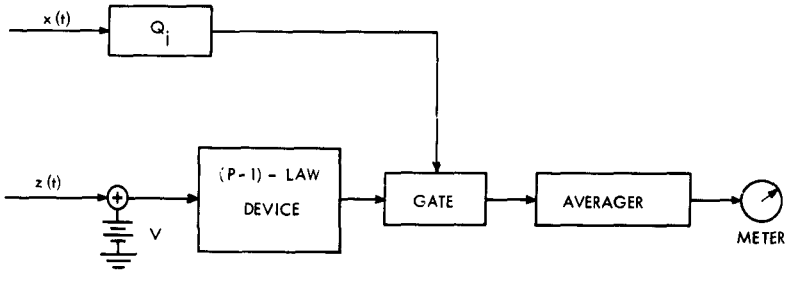


Fig. IX-3. A circuit for determining a_j for P even.

If P is odd, then there are two basically different circuits by which the desired value of a_j can be determined. One is obtained by noting that

$$|z-a_j|^{P-1}\text{Sgn}(z-a_j) = (z-a_j)^{P-1}|z-a_j| \qquad [P \geq 3 \text{ and odd}]. \tag{26}$$

A circuit, based on Eq. 26, for experimentally determining the desired value of a_j is depicted in Fig. IX-4. The second circuit is obtained by noting that a system whose output is Sgn $(z-a_j)$ is a saturating amplifier. Thus, for P odd, a second circuit is shown in Fig. IX-5. We note that the multiplier in Fig. IX-5 need only be a polarity-reversing switch that is controlled by the output of the saturating amplifier. In the special case for P = 1, the coefficients for the system with a minimum mean magnitude

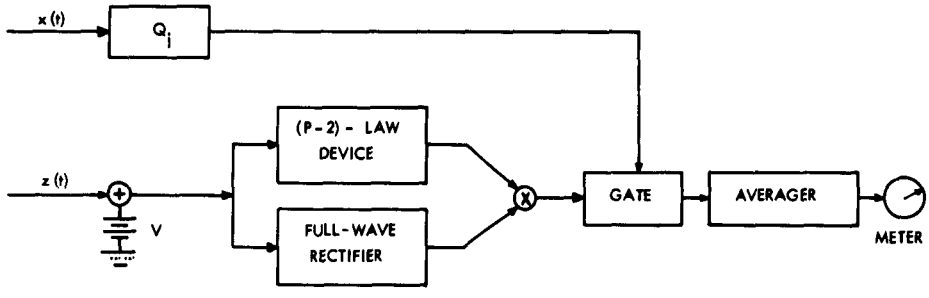


Fig. IX-4. A circuit for determining a_j for $P \geq 3$ and odd.

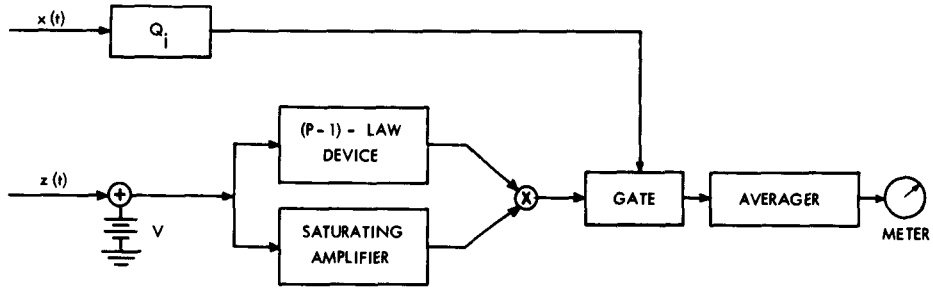


Fig. IX-5. A circuit for determining a_j for P odd.

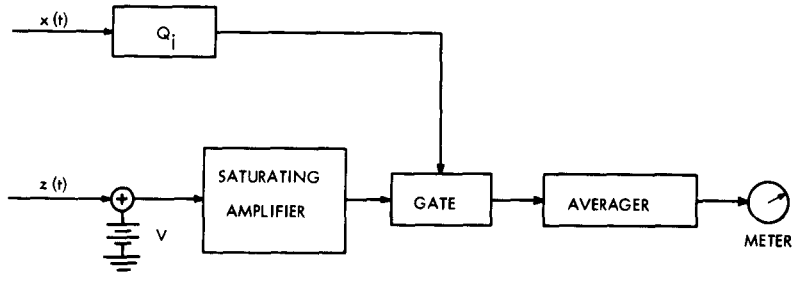


Fig. IX-6. A circuit for determining a_j for minimum $\overline{|E(t)|}$.

error are obtained. The circuit for determining these coefficients is relatively simple and is depicted in Fig. IX-6.

3. Optimum No-Memory Systems for Minimum $\overline{F[E(t)]}$

The procedure that we have just described can be extended to the determination of optimum nonlinear no-memory systems for error criteria that are arbitrary functions of the error. To extend this procedure, we have from Eq. 10

$$F[E] = F\left[z - \sum_{n=0}^{N} a_n Q_n(x)\right]. \tag{27}$$

According to the definition of the gate functions (Eqs. 3), at any instant only one term in the sum is nonzero. If, at that instant, x(t) is in the interval of the j^{th} gate function, then, at that instant, $F(E) = (z-a_j)\,Q_j(x)$. Thus, by the use of Eq. 5, we can express Eq. 27 as

$$F[E] = \sum_{n=0}^{N} F[z-a_n]\,Q_n(x). \tag{28}$$

Equation 28 is identical with Eq. 15 for $F(E) = E^P$. The set of coefficients for which the mean of Eq. 28 is stationary now can be determined by setting the derivative of $\overline{F(E)}$ with respect to a_j equal to zero. Thus, if we define

$$G(a) = \frac{d}{da} F(a), \tag{29}$$

then the condition that a_j be an extremal of the mean of Eq. 28 is

$$\overline{G(z-a_j)\,Q_j(x)} = 0. \tag{30}$$

A sufficient condition that assures us that the condition given by Eq. 30 yields a minimum of $\overline{F(E)}$ is that

$$\frac{d^2}{da^2} F(a) \geq 0. \tag{31}$$

Equation 31 is the statement that the error criterion is a convex function of its argument. For such cases, a circuit for experimentally determining the value of a_j that satisfies Eq. 30 is depicted in Fig. IX-7. The battery voltage, V, in the figure is adjusted to make the meter read zero; V is then numerically equal to the optimum value of a_j. If F(E) is not a convex function of its argument, this simple procedure is not sufficient, for local minima and maxima can then exist. However, the adjustment of any one coefficient affects only one term of the sum in Eq. 28. The minimum

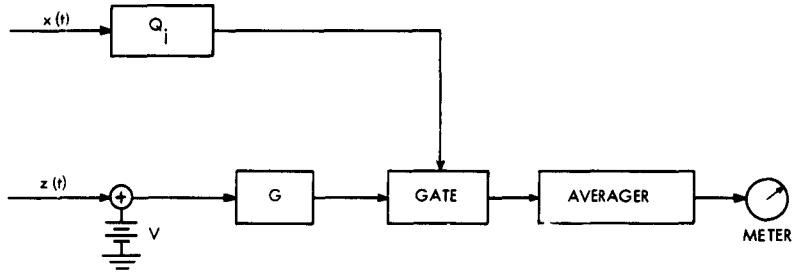


Fig. IX-7. A circuit for determining a_j for minimum $\overline{F[E(t)]}$.

of F(E) is thus obtained only if each term in Eq. 28 is a minimum. Therefore, because of the special properties of the gate functions, the amplifier gains, a_n, do not interact even for arbitrary functions of the error and thus they can be determined individually. This affords enormous simplification in search procedures for the determination of optimum systems for nonconvex functions of the error. However, if $G(a)$ can be expressed as a polynomial, then the values of a_j which satisfy Eq. 30 can be determined without a search procedure. For, if $G(a)$ can be expressed as a polynomial, then from Eq. 30 a polynomial expression in a_j can be obtained. The coefficients of the various powers of a_j will involve only the averages $\overline{z^n Q_j(x)}$ which can be measured experimentally. The values of a_j which satisfy Eq. 30 can then be determined by solving for the roots of the polynomial in a_j. Thus, we have the important result that for the cases in which F(E) can be expressed as a polynomial, no search procedure is required; then the optimum value of a_j can be determined analytically by solving for the minimum of the polynomial, $\overline{F(z-a_j)\, Q_j(x)}$.

4. An Example

An important consequence of the fact that the amplifier gains do not interact is that, in many cases, analytical estimates of the desired optimum nonlinear system can be obtained. We shall consider a simple example as an illustration. A received signal is the sum of a message and a noise that are statistically independent. The optimum nonlinear no-memory system is to be determined for an output that is the message. The error criterion to be used is $\overline{|E|^P}$ in which $P > 1$, but not necessarily an integer. The message is a binary signal that assumes the value 1 with probability 1/2 or the value -1 with probability 1/2. We shall determine the coefficients of the optimum system from Eq. 30. To do this, let the random variable of amplitude of the message be ζ which takes on values z; let the random variable of amplitude of the noise be η which takes on values y. Then, the random variable of amplitude of the received signal, ξ, is

$$\xi = \zeta + \eta. \tag{32}$$

On the ensemble basis, Eq. 30 for our example is

$$\overline{G(\zeta - a_j)\, Q_j(\xi)} = 0. \tag{33}$$

For our problem, $F(\alpha) = |\alpha|^P$ so that from Eq. 29

$$G(\alpha) = P|\alpha|^{P-1}\mathrm{Sgn}(\alpha). \tag{34}$$

Thus Eq. 33 becomes

$$0 = \overline{|\zeta - a_j|^{P-1}\mathrm{Sgn}(\zeta - a_j)\, Q_j(\zeta+\eta)}$$

$$= \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz\, |z - a_j|^{P-1}\mathrm{Sgn}(z - a_j)\, Q_j(z+y)\, P_{\zeta,\eta}(z,y). \tag{35}$$

However, since the message is binary and independent of the noise, we have

$$P_{\zeta,\eta}(z,y) = P_{\zeta}(z)\, P_{\eta}(y)$$

$$= \frac{1}{2}[u(z-1)+u(z+1)]\, P_{\eta}(y), \tag{36}$$

in which u(z) is the unit impulse function. Substituting Eq. 36 in Eq. 35 and integrating with respect to z, we obtain

$$0 = \frac{1}{2}|1-a_j|^{P-1}\mathrm{Sgn}(1-a_j) \int_{-\infty}^{\infty} Q_j(y+1)\, P_{\eta}(y)\, dy$$

$$+ \frac{1}{2}|-1-a_j|^{P-1}\mathrm{Sgn}(-1-a_j) \int_{-\infty}^{\infty} Q_j(y-1)\, P_{\eta}(y)\, dy. \tag{37}$$

The equation that a_j must satisfy is thus

$$-\left|\frac{1-a_j}{1+a_j}\right|^{P-1} \frac{\mathrm{Sgn}(1-a_j)}{\mathrm{Sgn}(-1-a_j)} = \frac{\int_{-\infty}^{\infty} Q_j(y-1)\, P_{\eta}(y)\, dy}{\int_{-\infty}^{\infty} Q_j(y+1)\, P_{\eta}(y)\, dy}. \tag{38}$$

We note that the right-hand side of the equation is the ratio of the probability that $(\eta-1)$ is in the interval of the j^{th} gate function to the probability that $(\eta+1)$ is in the interval of the j^{th} gate function. Let us denote this ratio, which is positive, by β_j. We then have

$$-\left|\frac{1-a_j}{1+a_j}\right|^{P-1} \frac{\mathrm{Sgn}(1-a_j)}{\mathrm{Sgn}(-1-a_j)} = \beta_j. \tag{39}$$

Since $\beta \geq 0$, we require for a solution that

$$\frac{\mathrm{Sgn}(1-a_j)}{\mathrm{Sgn}(-1-a_j)} < 0.$$

Thus we require that $|a_j|$ be less than one. Equation 39 can thus be written as

$$\left(\frac{1-a_j}{1+a_j}\right)^{P-1} = \beta_j; \qquad |a_j| < 1. \tag{40}$$

For any set of gate functions, β_j can be determined; a_j can then be obtained by means of Eq. 40. If each gate function is chosen to be of infinitesimal width, then

$$\beta_j = \beta(y) = \frac{P_\eta(y-1)}{P_\eta(y+1)}, \tag{41}$$

in which $P_\eta(y)$ is the probability density of the amplitude of the noise. For such a case, $a_j = a(y)$ is the transfer characteristic of the no-memory system. Solving for a(y) from Eq. 40, we then have

$$a(y) = \frac{1 - [\beta(y)]^{1/(P-1)}}{1 + [\beta(y)]^{1/(P-1)}}; \qquad P > 1 \tag{42}$$

in which $\beta(y)$ is given by Eq. 41. We thus have obtained an explicit expression for the transfer characteristic of the no-memory system. The analogy between the approach taken in this example and the solution of a differential equation by means of difference equations is to be noted.

5. Probability of a Function of the Error

Since an arbitrary function of the error can be written in the form of Eq. 28, we can also determine the coefficients to minimize Prob $\{F[E] > A\}$. We have from Eq. 28 the condition

$$\mathrm{Prob}\left\{\sum_{n=0}^{N} F[z-a_n]\, Q_n(x) > A\right\} = \text{minimum}. \tag{43}$$

But since, at any instant, only one term in the sum is nonzero, Eq. 43 can be written

$$\sum_{n=0}^{N} \mathrm{Prob}\left\{F[z-a_n]\, Q_n(x) > A\right\} = \text{minimum}. \tag{44}$$

However, the adjustment of any one coefficient affects only one term of the sum in Eq. 44. The minimum of Eq. 44 is thus obtained only if each term of the sum is a minimum. Thus the optimum value of a_j is that for which

$$\text{Prob}\{F[z-a_j]\, Q_j(x) > A\} = \text{minimum}. \tag{45}$$

A circuit by means of which the optimum value of a_j can be determined is depicted in Fig. IX-8. The voltage, V, is adjusted to minimize the meter reading; V is then

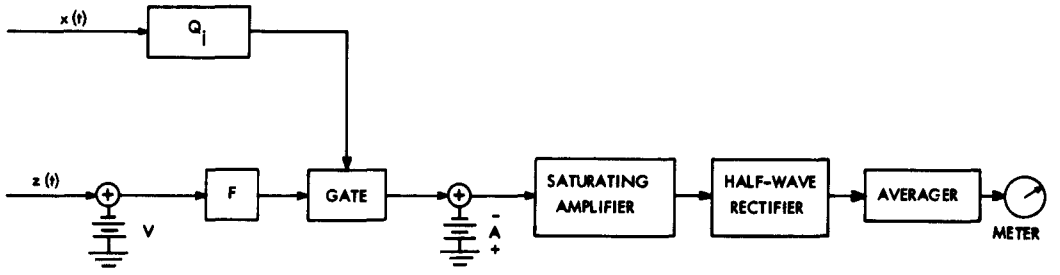


Fig. IX-8. A circuit for determining a_j that satisfies Eq. 44.

numerically equal to the desired value of a_j and the meter reading is the probability that $F[z-a_j]\, Q_j(x) > A$.

6. Weighted-Mean Function of the Error

All of the results that we have obtained can be extended to weighted functions of the error. By a weighted function of the error we mean $W(t)\, F[E(t)]$ in which $W(t)$ is an arbitrary function of time. For example, $W(t)$ can be a function of $[z(t)-x(t)]$. By use of Eq. 28, the weighted function of the error can be written

$$WF[E] = \sum_{n=0}^{N} F[z-a_n]\, WQ_n(x). \tag{46}$$

Thus all of our results apply if we replace $Q_n(x)$ by $WQ_n(x)$. Thus, to minimize a weighted mean of a convex function of the error, we have from Eq. 30 the result that the optimum value of a_j is that for which

$$\overline{G[z-a_j]\, WQ_j(x)} = 0. \tag{47}$$

7. Systems with Memory

We shall now present an extension of our previous results to nonlinear systems with memory. A complete set of operators that could be used for the synthesis of nonlinear

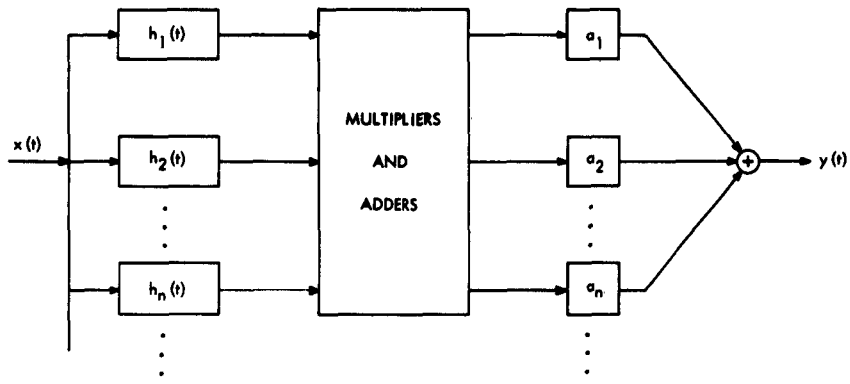


Fig. IX-9. Nonlinear system.

systems with memory is Wiener's G-functionals.[4] These operators can be synthesized in the form shown schematically in Fig. IX-9. As shown, the nonlinear system can be divided into three sections that are connected in tandem: a linear system whose impulse responses, $h_n(t)$, form a complete set of functions; a nonlinear no-memory section that consists of just multipliers and adders; and a section consisting of amplifiers whose outputs are then summed. In this model of a nonlinear system, the first section is the only one that has memory, since its outputs, $v_n(t)$, are the result of linear operations on the past of the input, x(t). We note from this model that the output, y(t), of a nonlinear system of the Wiener class is just a linear combination of various products of the outputs, $v_n(t)$, of the linear section. If the linear section consists of only K linear systems, then, according to the model, the output, y(t), can be expressed as

$$y = \sum_{k=1}^{\infty} \sum_{i_1=1}^{K} \cdots \sum_{i_k=1}^{K} A_{i_1,\ldots,i_k} v_{i_1} \cdots v_{i_k}. \tag{48}$$

By use of the gate functions, any step approximation to $v_i(t)$ can be given by

$$v_i \approx \sum_{n=0}^{N} B_n Q_n[v_i]. \tag{49}$$

By substituting Eq. 49 in Eq. 48, we note that we can express any step approximation to y(t) as

$$y \approx \sum_{k=1}^{K} \sum_{n_1=0}^{N} \cdots \sum_{n_k=0}^{N} C_{n_1,\ldots,n_k} Q_{n_1}(v_1) \ldots Q_{n_k}(v_k). \tag{50}$$

We note that by use of Eq. 5 we can write

$$\sum_{n_{k+1}=0}^{N} \cdots \sum_{n_K=0}^{N} Q_{n_{k+1}}(v_{k+1}) \cdots Q_{n_K}(v_K) = 1. \tag{51}$$

Thus Eq. 50 can be written in the form

$$y \approx \sum_{n_1=0}^{N} \cdots \sum_{n_K=0}^{N} D_{n_1,\ldots,n_K} Q_{n_1}(v_1) \cdots Q_{n_K}(v_K). \tag{52}$$

Define

$$\Phi_\alpha(\underline{v}) = Q_{n_1}(v_1) \cdots Q_{n_K}(v_K), \tag{53}$$

in which $\alpha = (n_1, n_2, \ldots, n_K)$. Then Eq. 52 can be written in the form

$$y \approx \sum_{\alpha} D_\alpha \Phi_\alpha(\underline{v}). \tag{54}$$

We note that the functions, $\Phi_\alpha(\underline{v})$, are K-dimensional gate functions, since $\Phi_\alpha(\underline{v})$ is nonzero and equal to one only when the amplitude of $v_1(t)$ is in the interval of $Q_{n_1}(v_1)$, and the amplitude of $v_2(t)$ is in the interval of $Q_{n_2}(v_2)$, and so on for each $v_n(t)$. That is, $\Phi_\alpha(\underline{v})$ is nonzero and equal to one only when $\underline{v}$ is in the α^{th} cell. We also note that

$$\sum_{\alpha} \Phi_\alpha(\underline{v}) = 1. \tag{55}$$

By use of the K-dimensional gate functions, all of our results can be extended to nonlinear systems that have memory. For we note that if the desired output is z(t), then any function of the error is

$$F[E] = F\left[z - \sum_{\alpha} A_\alpha \Phi_\alpha(\underline{v})\right]. \tag{56}$$

According to the definition of the K-dimensional gate functions, only one term in the sum is nonzero at any instant. If, at that instant, $\underline{v}$ is in the α^{th} cell, then at that instant $F[E] = F(z-A_\alpha)\, \Phi_\alpha(\underline{v})$. Thus, by use of Eq. 55, we can express Eq. 56 as

$$F[E] = \sum_{\alpha} F[z-A_\alpha]\, \Phi_\alpha(\underline{v}). \tag{57}$$

This equation is identical in form with Eq. 28 and thus all of the results that we have obtained for the one-dimensional case also apply to the K-dimensional case. For example, from Eq. 30 the condition that A_a be an extremal of $\overline{F(E)}$ is

$$\overline{G[z-A_a]\,\Phi_a(\underline{v})} = 0. \tag{58}$$

Experimentally, the values of A_a that satisfy Eq. 58 can be obtained by means of the circuit depicted in Fig. IX-10. The procedure is to adjust the battery voltage, V, until the meter reads zero; V is then numerically equal to the desired value of A_a for which

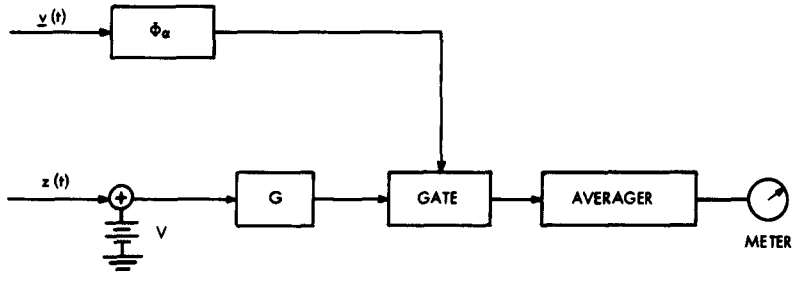


Fig. IX-10. A circuit for determining A_a for which $\overline{F[E]}$ is a minimum.

Eq. 58 is satisfied. However, if G(a) can be expressed as a polynomial, then a polynomial expression in A_a can be obtained from Eq. 58. The coefficients of the various powers of A_a will involve only the averages $\overline{z^n \Phi_a(\underline{v})}$ which can be measured experimentally. The values of A_a that satisfy Eq. 58 can then be determined by solving for the roots of the polynomial. For example, if $F(E) = E^2$, then Eq. 58 becomes

$$\overline{(z-A_a)\,\Phi_a(\underline{v})} = 0. \tag{59}$$

Solving for A_a, we have

$$A_a = \frac{\overline{z\Phi_a(\underline{v})}}{\overline{\Phi_a(\underline{v})}}. \tag{60}$$

Equation 60 is the result obtained by Bose.[3] We thus have the important result that, as in the case of no-memory systems, no adjustment procedure of any sort is required in those cases in which F(E) can be expressed as a polynomial. If F(E) is not a convex function, the optimum value of A_a can still be determined analytically by solving for the minimum of the polynomial, $\overline{F[z-A_a]\,\Phi_a(\underline{v})}$, and thus no search procedure is required. Also, the set of coefficients $\{A_a\}$ can be determined for which the mean of

an arbitrary weighted function of the error or the probability of an arbitrary function of the error is a minimum. Each coefficient of this set can be determined individually in a manner similar to that used for the no-memory case.

M. Schetzen

References

1. M. Schetzen, Some Problems in Nonlinear Theory, Technical Report 390, Research Laboratory of Electronics, M.I.T., July 6, 1962.

2. D. Sakrison, Application of Stochastic Approximation Methods to System Optimization, Technical Report 391, Research Laboratory of Electronics, M.I.T., July 10, 1962.

3. A. Bose, A Theory of Nonlinear Systems, Technical Report 309, Research Laboratory of Electronics, M.I.T., May 15, 1956.

4. N. Wiener, Nonlinear Problems in Random Theory (The Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1958).

B. OPTIMUM QUANTIZATION FOR A GENERAL ERROR CRITERION

1. Introduction

Quantization is the nonlinear, no-memory operation of converting a continuous signal to a discrete signal that assumes only a finite number of levels (N). Quantization occurs whenever it is necessary to represent physical quantities numerically. The primary concern in quantization is faithful reproduction, with respect to some fidelity criterion, of the input at the output. Thus, it is not necessary to require uniform quantization. In fact, since the number of output levels, N, will be specified, the "error" in the output will be minimized by adjusting the quantizer characteristic. Figure IX-11 illustrates the input-output characteristic of a quantizer.

Early investigations into the process of quantization considered speech to be the quantizer input signal. One of the first investigators was Bennett[1] who concluded that, with speech, it would be advantageous to taper the steps of the quantizer in such a manner that finer steps would be available for weak signals. Following Bennett, other investigators such as Smith,[2] Lozovoy,[3] and Lloyd[4] worked toward the characterization of optimum quantizers by assuming that a large number of levels would be used in the quantizer. Max[5] formulated the general problem of selecting the parameters of the optimum quantizer for a wide class of error criteria irrespective of the number of levels in the device. He also was able to determine a partial solution to the problem by paying particular attention to the mean-square error criteria. Bluestein[6] derived some extensions to Max's work for the special case of the mean-absolute error criteria.

In this report, the expression for the quantization error as a function of the quantizer

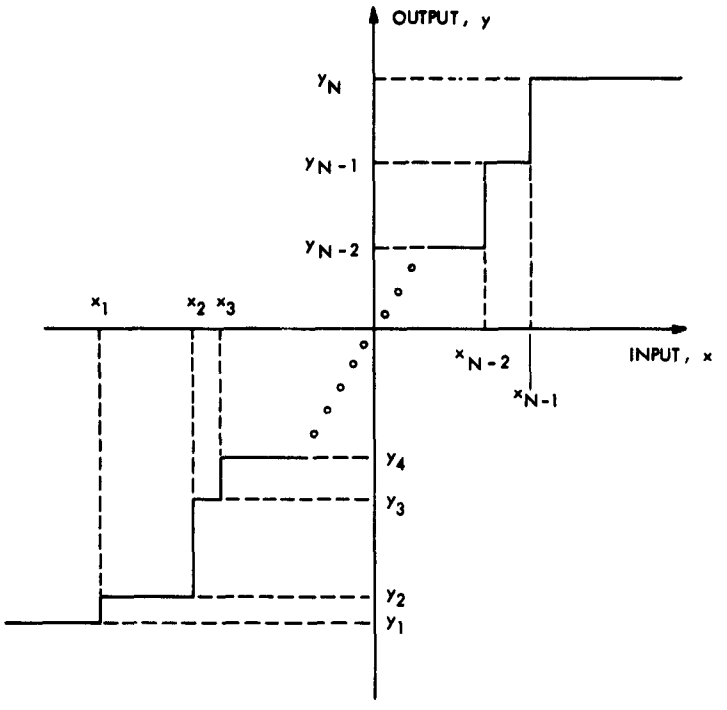


Fig. IX-11. Input-output relationship of the N-level quantizer.

parameters is presented and a method for determining these parameters by utilizing the technique of dynamic programming is developed. This method has two advantages over the previous ones. First, the method "searches" for the absolute minimum within the region of variation rather than the relative minima. Second, the method is guaranteed to converge to the absolute minimum in a specified number of steps.

2. Formulation of the Problem

Figure IX-11 shows the input-output relationship for the general quantizer. The output is y_k for $x_{k-1} \leq x < x_k$. The x_k are called the transition values; that is, x_k is the value of the input variable x for which there is a transition in the output from y_k to y_{k+1}. The y_k are called the representation values.

In the general quantizer problem under consideration here, the x_i, as well as the y_k (a total of 2N-1 quantities), are variables. These 2N-1 variables are to be chosen in such a manner that the error with respect to a specified error criterion is minimized.

In order to determine the specific values of x_i and y_k which specify the optimum quantizer, that is, the quantities that we shall call X_i and Y_k, we must first derive an

expression for the quantization error criterion. Since by optimum we shall mean minimum quantization error with respect to the specified error criterion, the X_i and the Y_k are the specific values of x_i and y_k which yield the absolute minimum value for the quantization error.

There is one set of constraints which must be imposed upon solutions to this problem. For purposes of organization as indicated in Fig. IX-11 we shall require that

$$\left.\begin{aligned} x_1 &< x_2 \\ x_2 &< x_3 \\ &\vdots \\ x_{N-2} &< x_{N-1} \end{aligned}\right\}. \tag{1}$$

With respect to the quantizer of Fig. IX-11, the error signal when the input signal satisfies the inequality

$$x_i \leq x < x_{i+1} \tag{2}$$

is

$$x - y_{i+1}. \tag{3}$$

We shall choose the error criterion to be the expected value of the function

$$g(x,y_{i+1}).$$

Then the measure of the error in this interval is

$$\int_{x_i}^{x_{i+1}} g(x,y_{i+1})\, p(x)\, dx, \tag{4}$$

where p(x) is the amplitude probability density of the input variable x. Since the error in each of the quantization intervals is independent of the error in the other intervals, the total quantization error is

$$\mathcal{E}(x_1, x_2, \ldots, x_{N-1}; y_1, y_2, \ldots, y_N) = \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} g(x,y_{i+1})\, p(x)\, dx. \tag{5}$$

In writing this equation we included the two parameters x_o and x_N for convenience. If x is a bounded function with lower bound X_ℓ and upper bound X_u, then, by definition, x_o will equal X_ℓ and x_N will equal X_u.

Upon first observation it appears that the coordinates of the absolute minimum value of (5) can be determined by use of the methods of calculus. Indeed, if there is only a single critical point of Eq. 5 within the region of variation (the region specified by Eq. 1)

and if this critical point is a relative minimum, then it is also the absolute minimum of the function. However, if there is more than a single critical point within the region of variation, then one of these critical points may be the absolute minimum of (5). Or it might be that none of these critical points is the absolute minimum. Thus the method of calculus is not a satisfactory technique for solving this problem. What is needed is a method that yields the coordinates of the absolute minimum whether or not the absolute minimum is at a critical point. The method of dynamic programming[7,8] is such a technique.

3. Determining the Optimum Quantizer by Using Dynamic Programming

In order to determine the coordinates of the absolute minimum value of (5) it is necessary to define a sequence of error functionals

$$\{f_i(x_i)\} \qquad i = 1, 2, \ldots, N$$

as follows.

$$\left.\begin{aligned}
f_1(x_1) &= \min_{\substack{y_1 \\ X_\ell \le x_1 \le X_u}} \left[\int_{X_\ell}^{x_1} g(x,y_1)\, p(x)\, dx\right] \\
f_2(x_2) &= \min_{\substack{y_2, x_1 \\ X_\ell \le x_1 \le x_2 \le X_u}} \left[\int_{x_1}^{x_2} g(x,y_2)\, p(x)\, dx + f_1(x_1)\right] \\
f_3(x_3) &= \min_{\substack{y_3, x_2 \\ X_\ell \le x_2 \le x_3 \le X_u}} \left[\int_{x_2}^{x_3} g(x,y_3)\, p(x)\, dx + f_2(x_2)\right] \\
&\vdots \\
f_i(x_i) &= \min_{\substack{y_i, x_{i-1} \\ X_\ell \le x_{i-1} \le x_i \le X_u}} \left[\int_{x_{i-1}}^{x_i} g(x,y_i)\, p(x)\, dx + f_{i-1}(x_{i-1})\right] \\
&\vdots \\
f_{N-1}(x_{N-1}) &= \min_{\substack{y_{N-1}, x_{N-2} \\ X_\ell \le x_{N-2} \le x_{N-1} \le X_u}} \left[\int_{x_{N-2}}^{x_{N-1}} g(x,y_{N-1})\, p(x)\, dx + f_{N-2}(x_{N-2})\right] \\
f_N(x_N) &= \min_{\substack{y_N, x_{N-1} \\ X_\ell \le x_{N-1} \le x_N \le X_u}} \left[\int_{x_{N-1}}^{x_N} g(x,y_N)\, p(x)\, dx + f_{N-1}(x_{N-1})\right]
\end{aligned}\right\} \quad (6)$$

It should be noted that in the last member of Eq. 6 we have taken the liberty of permitting x_N to take on variation from X_ℓ to X_u in deference to our previous assumption (see explanation immediately following Eq. 5) that $x_N = X_u$. This variation is necessary if $f_N(x_N)$ is to be completely determined.

Also, from Eq. 6 we can show that the last member of this set of equations, when evaluated at $x_N = X_u$, can be written

$$f_N(X_u) = \min_{\substack{y_i, x_{i-1} \\ i=1,2,\ldots N \\ X_\ell = x_0 \leq x_1 \leq \ldots \leq x_N = X_u}} \left[\sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} g(x, y_{i+1})\, p(x)\, dx \right]. \tag{7}$$

Therefore, $f_N(X_u)$ is the error for the optimum quantizer with N levels. (Note that, in general, $f_i(X_u)$ is the error for the optimum quantizer with i levels, $i \leq N$.)

It is possible to simplify the minimization process used to determine the error functionals, Eq. 6, when we realize that the minimum, with respect to y_i, occurs for a value of y_i which satisfies the equation

$$0 = \int_{x_{i-1}}^{x_i} \frac{\partial}{\partial y_i} [g(x, y_i)]\, p(x)\, dx. \tag{8}$$

If, as will usually be the case, $g(x, y_i)$ is greater than or equal to zero and is concave upward, then (8) has only a single solution. This solution will be denoted by $\overline{y}_i$. (Should (8) have more than one solution, the solution that minimizes (6) will be called $\overline{y}_i$.) Equation 8 then allows us to rewrite the general term of (6) and to eliminate the formal minimization with respect to y_i, since the value of y_i which minimizes $f_i(x_i)$ for specific values of x_i and x_{i-1} is $\overline{y}_i$. Thus (6) becomes

$$f_i(x_i) = \min_{\substack{x_{i-1} \\ X_\ell \leq x_{i-1} \leq x_i \leq X_u}} \left[\int_{x_{i-1}}^{x_i} g(x, \overline{y}_i)\, p(x)\, dx + f_{i-1}(x_{i-1}) \right] \qquad i = 1, 2, \ldots, N \tag{9}$$

where $x_0 = X_\ell$, a constant, and $f_0(X_\ell) \equiv 0$. With reference to Eq. 7 we see that (8) has reduced the number of variables over which the formal minimization must be performed from 2N-1 to N.

In order to determine the optimum quantizer parameters it is necessary to define two auxiliary sets of functionals; first, the transition-value decision functionals

$$\{X_i(x)\} \qquad i = 1, 2, \ldots, N;$$

and second, the representation-value decision functionals

$$\{Y_i(x)\} \qquad i = 1, 2, \ldots, N.$$

The transition-value decision functionals are defined as follows

$$\left.\begin{aligned} X_1(x) &= x; \\ X_2(x) &= \text{the value of } x_1 \text{ which minimizes } f_2(x_2) \text{ for a specified } x = x_2,\ y_2 = \overline{y_2}; \\ &\vdots \\ X_N(x) &= \text{the value of } x_{N-1} \text{ which minimizes } f_N(x_N) \text{ for a specified } x = x_N,\ y_N = \overline{y_N} \end{aligned}\right\}. \tag{10}$$

In a similar manner the representation-value decision functionals are defined as

$$\left.\begin{aligned} Y_1(x) &= \overline{y_1} \text{ for a specified } x = x_1; \\ Y_2(x) &= \overline{y_2} \text{ when } x_1 \text{ is that value [i.e., } x_1 = X_2(x)] \text{ which minimizes } f_2(x_2) \text{ for a specified } x = x_2; \\ &\vdots \\ Y_N(x) &= \overline{y_N} \text{ when } x_{N-1} \text{ is that value [i.e., } x_{N-1} = X_N(x)] \text{ which minimizes } f_N(x_N) \text{ for a specified } x = x_N \end{aligned}\right\}. \tag{11}$$

Each of the functionals, Eqs. 6, 10, and 11, will be represented by a tabulation of its value at a number of points taken along an equispaced grid. In general, the same set of points will be used for each of the functionals.

As we have previously indicated, $f_N(X_u)$ is the measure of error for the optimum quantizer with N levels with the specified error criterion used. Therefore, the location of the transition value between levels (N-1) and (N) is specified by (10), that is,

$$X_{N-1} = X_N(X_u).$$

This level will be represented by the representation value

$$Y_N = Y_N(X_u),$$

specified by Eq. 11. At this point in the decision process, x that is such that $X_{N-1} \leq x \leq X_u$ has been allocated to the N^{th} level. The remaining values of x, $X_\ell \leq x \leq X_{N-1}$, remain to be quantized into N-1 levels. From (10) the transition value between levels (N-2) and (N-1) is given by

$$X_{N-2} = X_{N-1}(X_{N-1}),$$

and from (11) the representation value is

$$Y_{N-1} = Y_{N-1}(X_{N-1}).$$

This decision process is continued until the first representation value is found from (11) to be

$$Y_1 = Y_1(X_1).$$

Once the decision process is completed, the optimum quantizer with respect to the specified error criterion is determined.

As a first example, this method of selecting the optimum quantizer is being applied in the case in which speech is the input signal. In this application the error criterion will be the expected value of the function

$$g(x,y_{i+1}) = |x-y_{i+1}|^{\nu} W(x),$$

where

$$W(x) \geq 0 \qquad -\infty < x < \infty.$$

J. D. Bruce

References

1. W. R. Bennett, Spectra of quantized signals, Bell System Tech. J. 27, 446-472 (1948).

2. B. Smith, Instantaneous companding of quantized signals, Bell System Tech. J. 36, 653-709 (1957).

3. I. A. Lozovoy, Regarding the computation of the characteristics of compression in systems with pulse code modulation, Telecommunications (published by AIEE), No. 10, pp. 18-25, 1961.

4. S. P. Lloyd, Least Squares Quantization in PCM (unpublished manuscript, Bell Telephone Laboratories, Inc., 1958), reported by D. S. Ruchkin, Optimum Reconstruction of Sampled and Quantized Stochastic Signals, Doctor of Engineering Dissertation, Yale University, May 1960.

5. J. Max, Quantizing for minimum distortion, IRE Trans., Vol. IT-6, pp. 7-12, March 1960.

6. L. I. Bluestein, A Hierarchy of Quantizers, Ph.D. Thesis, Columbia University, May 1962.

7. R. Bellman and J. Dreyfus, Applied Dynamic Programming (Princeton University Press, Princeton, N. J., 1962).

8. R. Bellman and B. Kotkin, On the Approximation of Curves by Linear Segments Using Dynamic Programming—II, Memorandum RM-2978-PR, The Rand Corporation, Santa Monica, California, February 1962.

X. PROCESSING AND TRANSMISSION OF INFORMATION*

Prof. W. B. Davenport, Jr.
Prof. P. Elias
Prof. R. M. Fano
Prof. R. G. Gallager
Prof. F. C. Hennie III
Prof. E. M. Hofstetter
Prof. D. A. Huffman
Prof. I. M. Jacobs
Prof. A. M. Manders
Prof. B. Reiffen
Prof. W. F. Schreiber
Prof. C. E. Shannon
Prof. J. M. Wozencraft
Dr. C. L. Liu
Dr. A. Wojnar

T. G. Adcock
T. M. Anderson
P. G. Arnold
M. H. Bender
E. R. Berlekamp
D. G. Botha
J. E. Cunningham
H. Dym
P. M. Ebert
D. Ecklein
D. D. Falconer
E. F. Ferretti
G. D. Forney, Jr.
R. L. Greenspan

U. F. Gronemann
P. W. Hartman
A. R. Hassan
J. L. Holsinger
T. S. Huang
N. Imai
R. J. Purdy
J. F. Queenan
J. E. Savage
J. R. Sklar
I. G. Stiglitz
W. R. Sutherland
O. J. Tretiak
W. J. Wilson
H. L. Yudkin

A. TWO-DIMENSIONAL POWER DENSITY SPECTRUM OF TELEVISION RANDOM NOISE

1. Introduction

Several workers have investigated the subjective effect of television random noise.[1-4] They have tried to relate the subjective effect of additive Gaussian noise to the one-dimensional power density spectrum of the noise considered as a function of time. Although the noise is one-dimensionally generated as a function of time, it is nevertheless displayed in two-dimensional fashion. Therefore, it has been our opinion that it might be more fruitful to try to relate the subjective effect of additive Gaussian noise to the two-dimensional power density spectrum of the noise considered as a function of two space variables. In this connection the following question naturally arises. A one-dimensional Gaussian noise with known power density spectrum $\Phi_o(\omega)$ is given. From this noise a two-dimensional noise is obtained by scanning. What is the power density spectrum $\Phi^*(u,v)$ of the two-dimensional noise in terms of $\Phi_o(\omega)$? In this report we shall attempt to answer this question.

2. Two-Dimensional Power Density Spectrum

We restrict our attention to black-and-white still pictures. The two-dimensional noise is completely specified by its magnitude $n^*(x,y)$ as a function of the space variables x and y. The geometry of the scanning lines is shown in Fig. X-1. We assume that the scanning is from left to right and from top to bottom, and that there is no interlacing.

*This research was supported in part by Purchase Order DDL B-00368 with Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the joint support of the U.S. Army, Navy, and Air Force under Air Force Contract AF19(604)-7400; and in part by the National Institutes of Health (Grant MH-04737-02), and the National Science Foundation (Grant G-16526).

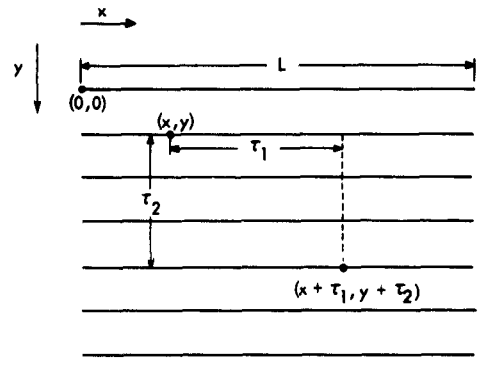


Fig. X-1. Geometry of scanning lines.

The length of each scan line is L, and the distance between two successive scan lines is taken as the unit length. Let the one-dimensional noise be $n_o(t)$, and let $(x,y) = (0,0)$ correspond to $t = 0$. Then we have

$$n^*(x,y) = n(x,y) \sum_k \delta(y-k), \tag{1}$$

where

$$n(x,y) = n_o(x+yL) \tag{2}$$

and

$$k = 0, \pm 1, \pm 2, \ldots; \quad -\infty < x < \infty, \quad -\infty < y < \infty.$$

Any particular noise picture can be considered as a finite piece of the sample function (1) which is infinite in extent. The impulses $\delta(y-k)$ are used so that we can deal with the Fourier transform of n^* instead of the Z-transform.

We assume that the one-dimensional noise is ergodic, then n(x,y) is also ergodic. The autocorrelation function of n(x,y) is

$$\begin{aligned}\phi(\tau_1,\tau_2) &= \overline{n(x,y)\, n(x+\tau_1, y+\tau_2)} \\ &= \overline{n(t)\, n(t+\tau_1+\tau_2 L)} \\ &= \phi_o(\tau_1+\tau_2 L),\end{aligned} \tag{3}$$

where $\phi_o(\tau)$ is the autocorrelation function of $n_o(t)$. Let $\Phi(u,v)$ and $\Phi_o(\omega)$ be the power density spectra of n(x,y) and $n_o(t)$, respectively. We have

$$\Phi(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(\tau_1,\tau_2)\, e^{j(\tau_1 u+\tau_2 v)}\, d\tau_1 d\tau_2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_o(\tau_1+\tau_2 L)\, e^{j(\tau_1 u+\tau_2 v)}\, d\tau_1 d\tau_2$$

$$= \frac{2\pi}{L}\,\delta\left(u-\frac{v}{L}\right)\,\Phi_o\left(\frac{v}{L}\right). \qquad (4)$$

Figure X-2 shows a possible $\Phi_o(\omega)$ and its corresponding $\Phi(u,v)$. It is to be noted that $\Phi(u,v)$ is zero everywhere except on the line $u-\frac{v}{L}=0$ where it is an impulse sheet.

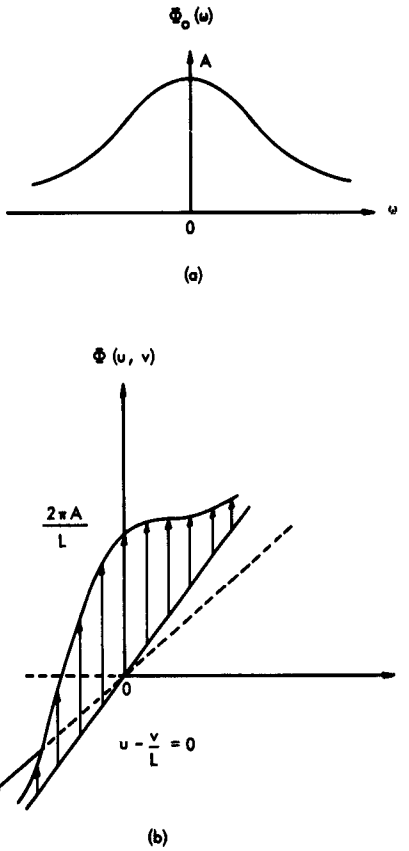


Fig. X-2. Relation between $\Phi(u,v)$ and $\Phi_o(\omega)$.

For ordinary commercial television, we have $L \approx 500$. Hence the line $u-\frac{v}{L}=0$ is very close to the v-axis.

Letting $\Phi^*(u,v)$ be the power density spectrum of $n^*(x,y)$, we have

$$\Phi^*(u,v) = \sum_k \Phi(u, v+2\pi k), \tag{5}$$

where the summation is over all integers. $\Phi^*(u,v)$ consists of identical impulse sheets on the lines $u - \frac{v}{L} = \frac{2\pi k}{L}$, $k = 0, \pm 1, \pm 2, \ldots$.

3. One-Dimensional Power Density Spectrum along a Particular Direction

We have derived the power density spectrum $\Phi^*(u,v)$ of a two-dimensional noise in terms of the one-dimensional power density spectrum along the x-direction. It is reasonable to believe that the subjective effect of noise depends also on the one-dimensional power density spectra along directions other than the x-axis.

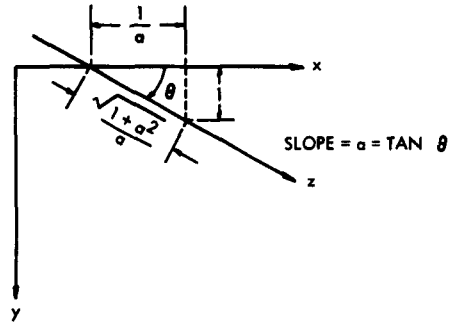


Fig. X-3. Calculation of the one-dimensional power density spectrum along the a direction.

We shall find the one-dimensional power density spectrum $\Phi_a^*(\omega)$ along a line of slope a (Fig. X-3). Let $n_a^*(z)$ be the noise along such a line:

$$n_a^*(z) = n_a(z) \sum_k \delta\left(z - \frac{\sqrt{1+a^2}}{a} k\right); \qquad k = 0, \pm 1, \pm 2, \ldots, \tag{6}$$

where

$$n_a(z) = n(z \cos\theta, z \cos\theta + b) \quad \text{for some fixed } b. \tag{7}$$

The autocorrelation function of $n_a(z)$ is

$$\phi_a(\tau) = \overline{n_a(z)\, n_a(z+\tau)}$$

$$= \overline{n(x,y)\, n(x+\tau\cos\theta,\, y+\tau\sin\theta)}$$

$$= \phi(\tau\cos\theta,\, \tau\sin\theta) \tag{8}$$

or

$$\phi'_a(\tau_1) \equiv \phi_a\left(\frac{\tau_1}{\cos\theta}\right) = \phi(\tau_1,\, a\tau_1). \tag{9}$$

The Fourier transform of $\phi'_a(\tau_1)$ is

$$\Phi'_a(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi(\omega - av, v)\, dv. \tag{10}$$

Hence the Fourier transform of $\phi_a(\tau_1)$ is

$$\Phi_a(\omega) = \frac{1}{|\cos\theta|}\,\Phi'_a\left(\frac{\omega}{\cos\theta}\right)$$

$$= \frac{1}{2\pi|\cos\theta|}\int_{-\infty}^{\infty} \Phi\left(\frac{\omega}{\cos\theta} - av, v\right) dv. \tag{11}$$

Putting Eq. 4 into Eq. 11, we have

$$\Phi_a(\omega) = \frac{1}{|(1+La)\cos\theta|}\,\Phi_o\left[\frac{\omega}{(1+La)\cos\theta}\right]$$

$$= \frac{\sqrt{1+a^2}}{|1+La|}\,\Phi_o\left(\frac{\sqrt{1+a^2}}{1+La}\,\omega\right). \tag{12}$$

In particular, for a = 0, we have

$$\Phi_o(\omega) = \Phi_o(\omega)$$

which checks with our assumption. For a = ∞, we have

$$\Phi_\infty(\omega) = \frac{1}{L}\,\Phi_o\left(\frac{\omega}{L}\right). \tag{13}$$

We note that for L ≈ 500, the bandwidth of $\Phi_\infty(\omega)$ is 500 times that of $\Phi_o(\omega)$. Figure X-4 shows how the factor $\frac{\sqrt{1+a^2}}{|1+La|}$ varies with the slope a. We note that the area under $\Phi_a(\omega)$ is independent of a.

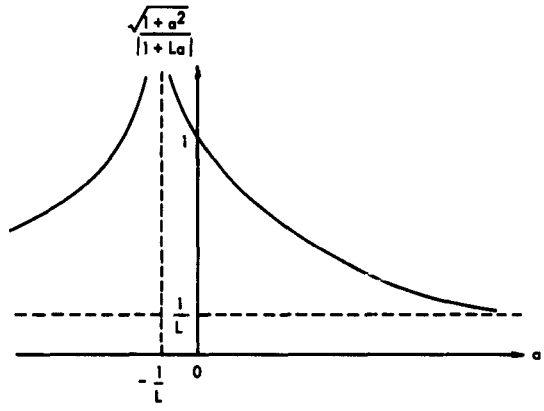


Fig. X-4. $\frac{\sqrt{1+a^2}}{|1+La|}$ vs a.

Finally, from Eq. 6, we find that the one-dimensional power density spectrum along the direction a is

$$\Phi_a^*(\omega) = \sum_k \Phi_a\left(\omega + \frac{2\pi a}{\sqrt{1+a^2}}k\right), \tag{14}$$

where $k = 0, \pm 1, \pm 2, \ldots$.

4. Discussion

From subjective tests, it has been found[1-4] that for pictures that are more or less isotropic, low-frequency noise (low-frequency when considered as a time function) is in general more annoying. From Eq. 12 we know, however, that a two-dimensional noise obtained from a low-frequency one-dimensional noise by scanning may contain high frequencies along directions other than the x-axis. In particular, the bandwidth along the y-axis is approximately 500 times as wide as that along the x-axis. Since the spatial frequency response of the human eye has a high-frequency cutoff, we suspect that the following hypothesis might be true for isotropic pictures contaminated by additive Gaussian noise. The more anisotropic a noise is, the more annoying it will be. Work is being carried out to test, among other things, this hypothesis, but the results are still not conclusive.

It is important to note that the mathematical model from which we obtained the power density spectra is quite crude. In order to relate the spectra to the subjective effect of noise, modifications may have to be made to take care

of the finiteness of scanning aperture.

We should like to thank W. L. Black who offered many very helpful comments on this work.

T. S. Huang

References

1. P. Mertz, Perception of TV random noise, SMPTE J. 54, 8-34 (January 1950).

2. P. Mertz, Data on random-noise requirements for theater TV, SMPTE J. 57, 89-107 (August 1951).

3. J. M. Barstow and H. N. Christopher, The measurement of random monochrome video interference, Trans. AIEE, Vol. 73, Part 1, pp. 735-741, 1954.

4. R. C. Brainard, F. W. Kammerer, and E. G. Kimme, Estimation of the subjective effects of noise in low-resolution TV systems, IRE Trans. on Information Theory, Vol. IT-8, pp. 99-106, February 1962.

B. SEQUENTIAL DECODING FOR AN ERASURE CHANNEL WITH MEMORY

1. Introduction

It has been shown that sequential decoding is a computationally efficient technique for decoding with high reliability on memoryless (or constant) channels.[1] It is of interest to determine whether or not similar statements are true for channels with memory.

In order to gain insight into the problem of sequential decoding with memory, a simple channel model, a two-state Markov erasure channel, has been analyzed. The presence of memory is shown to increase the average number of computations above that required for a memoryless erasure channel with the same capacity. For the erasure channel, the effects of memory can be reduced by scrambling.

2. Channel Model

Assume as a channel model an erasure channel with an erasure pattern that is generated by a Markov process. Assume the process shown in Fig. X-5. Assume for

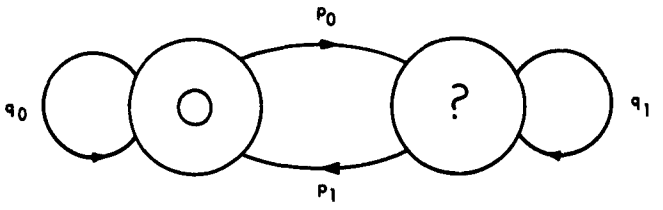


Fig. X-5. Noise model.

simplicity that the process begins in the 0 state. This will not affect the character of the results.

It can be shown that the probability of being in the 0 state after many transitions becomes independent of the starting probabilites and approaches $\frac{p_1}{p_0 + p_1}$.[2] Channel capacity is defined as

$$C = \lim_{n\to\infty} \frac{1}{n} \max_{p(x^n)} I(x^n;y^n),$$

where x^n and y^n are input and output sequences of n digits, respectively. Then, intuitively, $C = \frac{p_1}{p_0 + p_1}$, since information is transmitted only when the noise process is in the 0 state and the frequency of this event approaches $n\frac{p_1}{p_0 + p_1}$ with large n.

3. Decoding Algorithim

We encode for this channel in the following fashion: A stream of binary digits is supplied by the source. Each digit selects ℓ digits from a preassigned tree (see Fig. X-6). If a digit has the value 1, the upper link at a node is chosen, the lower link being chosen when it is 0. (In our example (1, 0, 0) produces (011, 101, 000).)

Our object in decoding will be to determine the first information digit of a sequence of n digits. Having done this, we move to the next node and repeat the decoding process, again using n digits. The following decoding algorithim is used: Build two identical decoders that operate separately on each subset of the tree in time synchronism. Each

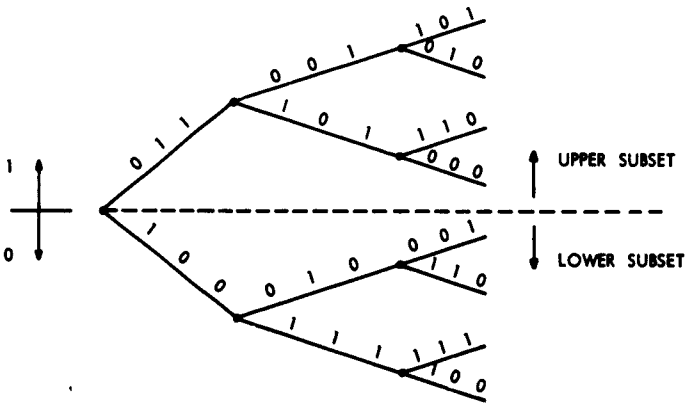


Fig. X-6. Encoding tree with $\ell = 3$.

decoder compares the received sequence with paths in its subset, discarding a path (and all longer paths with this as a prefix) as soon as a digit disagrees with an unerased received digit. When either a subset is identified (one subset is eliminated) or an ambiguity occurs (that is, when an acceptable path is found in each subject), we stop. If an ambiguity occurs, ask for a repeat. The computation, then, is within a factor of 2 of that necessary to decode the incorrect subset. We average the computation over the ensemble of all erasure patterns and all incorrect subsets.

4. Computation

Let x_i equal the number of nodes that are acceptable at the i^{th} state of the incorrect subset. We trace two links for each node or perform two computations. Then, the average computation to decode the incorrect subset is I(n).

$$I(n) = 2 \sum_{i=0}^{nR-1} \overline{x_i},$$

where $R = \frac{1}{\ell}$, the rate of the source. We have

$$\overline{x_i} = 2^{i-1} P_R \qquad \text{(a path of } i\ell \text{ digits is acceptable)}$$

$$= 2^{i-1} p_{i\ell}.$$

Now

$$p_m = \sum_{r=0}^{m} \left(\frac{1}{2}\right)^{m-r} P_m(r),$$

where $P_m(r)$ is the probability of r erasures in m transitions. Then

$$I(n) = \sum_{i=0}^{nR-1} \left(\frac{1}{2}\right)^{i(\ell-1)} \sum_{r=0}^{i\ell} 2^r P_{i\ell}(r).$$

We recognize that the sum on r is the moment-generating function of erasures, $g_m(s)$, evaluated at $s = \ln 2$.

$$g_m(s) = \sum_{e_m} e^{s\phi(e_m)} P(e_m),$$

where e_m is a sequence of m noise states and

$$\phi(e_m) = \sum_{i=1}^{m} \phi(e_i); \qquad \phi(x) = \begin{cases} 1 & x = ? \\ 0 & x \neq ? \end{cases}$$

It can be shown that

$$g_m(s) = \underline{1,\ 0}\,\Pi^{m-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

where $\Pi = \begin{bmatrix} q_o & p_o e^s \\ p_1 & q_1 e^s \end{bmatrix}$, and that Π^m is asymptotically equal to the m^{th} power of the largest eigenvalue of Π. Then, we find the following asymptotically correct expression for I(n)

$$I(n) \simeq \begin{cases} \omega_o & R < R_{comp} \\ \omega_1 2^{n(R-R_{comp})} & R > R_{comp} \end{cases}$$

where ω_o, ω_1 are constants and

$$R_{comp} = 1 - \log_2 \frac{1}{2}\left[3 - p_o\left(\frac{1+c}{1-c}\right) + \sqrt{\left(3-p_o\left(\frac{1+c}{1-c}\right)\right)^2 - 8\left(1 - \frac{p_o}{1-c}\right)^2}\right].$$

Using similar techniques, we can find an upper bound to the probability of ambiguity and show that it is equal to the random block-coding bound and that the zero rate exponent of this bound is R_{comp}.

The computational cutoff rate, R_{comp}, is shown as a function of p_o for fixed capacity, c, in Fig. X-7. R_{comp} is an increasing function of p_o for constant c. It is equal to R^o_{comp}, the memoryless rate, when $p_o = q_1 = 1 - c$, and it exceeds R^o_{comp} when $p_o > q_1$. However, $p_o > q_1$ is not, in general, physically meaningful, since this situation corresponds to anticlustering.

5. Conclusions

We conclude that sequential decoding is inefficient when the channel becomes "sticky" (small p_o). It is possible, however, to reduce the effective memory of the channel and increase R_{comp} if we scramble before transmission and unscramble after reception. Since the ergodic probability, $\frac{p_1}{p_o + p_1}$, of erasures is not changed by this maneuver, channel capacity remains constant.

Most of these results appear to be extendable to binary-input, binary-output channels. It is not clear, however, that scrambling will improve performance on binary channels,

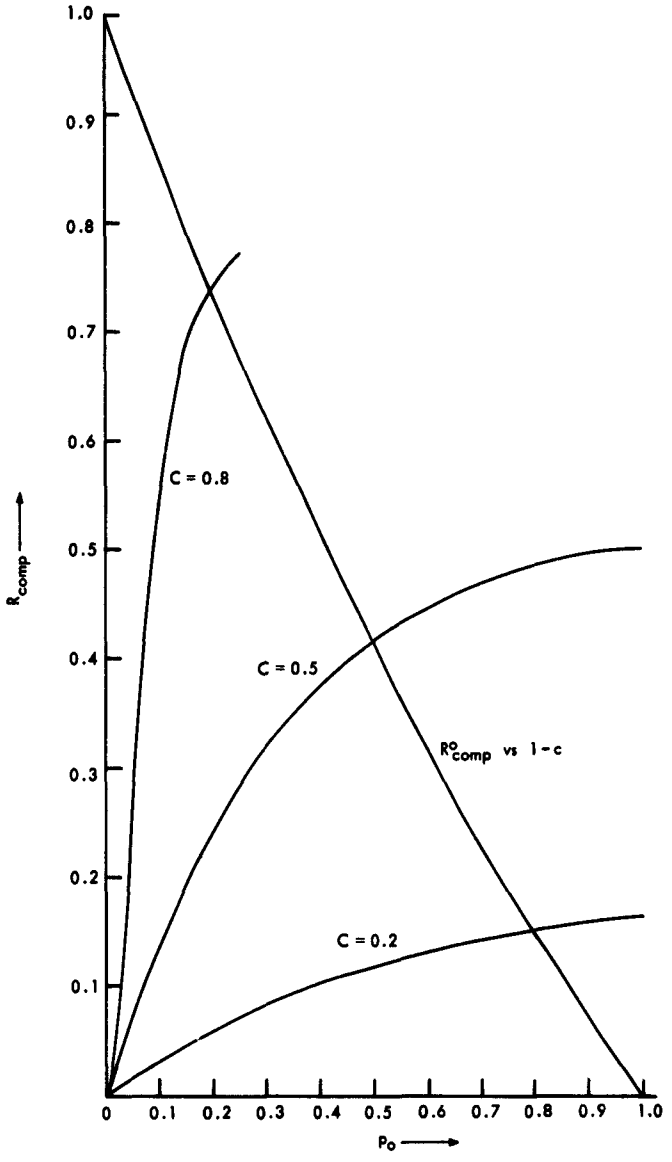


Fig. X-7. R_{comp} versus p_o.

since capacity is reduced in the process. Work is continuing on this problem.

J. E. Savage

References

1. J. M. Wozencraft and B. Reiffen, Sequential Decoding (The Technology Press of the Massachusetts Institute of Technology, Cambridge, Mass., 1961).

2. R. A. Howard, Dynamic Programming and Markov Processes (The Technology Press of the Massachusetts Institute of Technology, Cambridge, Mass., 1960), p. 7.

C. A SIMPLE DERIVATION OF THE CODING THEOREM

This report will summarize briefly a new derivation of the Coding Theorem. Let $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_M$ be a set of M code words for use on a noisy channel, and let $\underline{y}_1, \underline{y}_2, \cdots, \underline{y}_L$ be the set of possible received sequences at the channel output. The channel is defined by a set of transition probabilities on these sequences, $\Pr(\underline{y}|\underline{x}_m)$. Define

$$\Pr(\underline{y}) = \sum_m \Pr(\underline{x}_m) \Pr(\underline{y}|\underline{x}_m) \tag{1}$$

$$\Pr(\underline{x}_m|\underline{y}) = \frac{\Pr(\underline{y}|\underline{x}_m) \Pr(\underline{x}_m)}{\Pr(\underline{y})}. \tag{2}$$

For a given received $\underline{y}$, the decoder will select the number m for which $\Pr(\underline{x}_m|\underline{y})$ is a maximum. We call this number $m_{\underline{y}}$. The probability that the decoder selects correctly in this case is, then, $\Pr(\underline{x}_{m_{\underline{y}}}|\underline{y})$. Therefore the probability of decoding error is

$$P_e = \sum_{\underline{y}} \Pr(\underline{y}) \left[1 - \Pr\left(\underline{x}_{m_{\underline{y}}}|\underline{y}\right)\right]. \tag{3}$$

We now upper-bound the expression in brackets in Eq. 3.

$$1 - \Pr\left(\underline{x}_{m_{\underline{y}}}|\underline{y}\right) = \sum_{m \neq m_{\underline{y}}} \Pr(\underline{x}_m|\underline{y}) \tag{4}$$

$$\leq \left[\sum_{m \neq m_{\underline{y}}} \Pr(\underline{x}_m|\underline{y})^{1/(1+\rho)}\right]^{1+\rho} \qquad \text{for any } \rho > 0. \tag{5}$$

Equation 5 follows from the fact that $t^{1/(1+\rho)}$ is a convex upward function of t for $\rho \geq 0$. Rewriting Eq. 5, we obtain

$$1 - \Pr\left(\underline{x}_m \underset{\underline{y}}{} | \underline{y}\right) \leq \sum_{m' \neq m} \underset{\underline{y}}{} \Pr(\underline{x}|\underline{y})^{1/(1+\rho)} \left[\sum_{m' \neq m} \underset{\underline{y}}{} \Pr(\underline{x}_{m'}|\underline{y})^{1/(1+\rho)}\right]^{\rho} \tag{6}$$

$$\leq \sum_{m} \Pr(\underline{x}_m|\underline{y})^{1/(1+\rho)} \left[\sum_{m' \neq m} \Pr(\underline{x}_{m'}|\underline{y})^{1/(1+\rho)}\right]^{\rho}. \tag{7}$$

Equation 7 follows from overbounding the first term of Eq. 6 by summing over all m, and overbounding the second term by replacing the missing term, $\Pr\left(\underline{x}_m \underset{\underline{y}}{} | \underline{y}\right)$, with a smaller missing term, $\Pr(\underline{x}_m|\underline{y})$. Now assume that the code words are equally likely, so that $\Pr(\underline{x}_m) = \frac{1}{M}$ for all m, and substitute Eq. 2 in Eq. 7.

$$1 - \Pr\left(\underline{x}_m \underset{\underline{y}}{} | \underline{y}\right) \leq \frac{1}{M \Pr(\underline{y})} \sum_{m} \Pr(\underline{y}|\underline{x}_m)^{1/(1+\rho)} \left[\sum_{m' \neq m} \Pr(\underline{y}|\underline{x}_{m'})^{1/(1+\rho)}\right]^{\rho} \tag{8}$$

$$P_e \leq \frac{1}{M} \sum_{\underline{y}} \sum_{m} \Pr(\underline{y}|\underline{x}_m)^{1/(1+\rho)} \left[\sum_{m' \neq m} \Pr(\underline{y}|\underline{x}_{m'})^{1/(1+\rho)}\right]^{\rho}. \tag{9}$$

Equation 9 bounds P_e for a particular code in terms of an arbitrary parameter $\rho > 0$. We shall now average this result over an ensemble of codes. For each m, let $\underline{x}_m$ be chosen independently according to some probability measure $P(\underline{x})$.

$$\overline{P}_e \leq \frac{1}{M} \sum_{\underline{y}} \sum_{m} \sum_{\underline{x}_m} P(\underline{x}_m) \Pr(\underline{y}|\underline{x}_m)^{1/(1+\rho)} \left[\sum_{m' \neq m} \Pr(\underline{y}|\underline{x}_{m'})^{1/(1+\rho)}\right]^{\rho}. \tag{10}$$

The bar over the last term in Eq. 10 refers to the average over the ensemble of all code words other than m. Now let $\rho \leq 1$. Then we are averaging over a convex upward function, and we can upper-bound Eq. 10 by averaging before raising to the power ρ. This is, then, the average of a sum of M-1 identically distributed random variables. Thus, for $0 \leq \rho \leq 1$,

$$\overline{P}_e \leq \frac{1}{M} \sum_{\underline{y}} \sum_{m} \sum_{\underline{x}_m} P(\underline{x}_m) \Pr(\underline{y}|\underline{x}_m)^{1/(1+\rho)} \left\{(M-1) \sum_{\underline{x}} P(\underline{x}) \Pr(\underline{y}|\underline{x})^{1+\rho}\right\}^{\rho}. \tag{11}$$

Removing the index m in the sum over $\underline{x}_m$, and summing over m, we have

$$\overline{P}_e \leq \sum_{\underline{y}} \left[\sum_{\underline{x}} P(\underline{x}) \Pr(\underline{y}|\underline{x})^{1/(1+\rho)} \right]^{1+\rho} (M-1)^{\rho}. \tag{12}$$

The bound in Eq. 12 applies to any channel for which $\Pr(\underline{y}|\underline{x})$ can be defined, and is valid for all choices of $P(\underline{x})$ and all ρ, $0 \leq \rho \leq 1$. If the channel is memoryless (that is, if $\Pr(\underline{y}|\underline{x}) = \prod_{n=1}^{N} \Pr(y_n|x_n)$, where $\underline{y} = (y_1, \cdots y_n, \cdots y_N)$, $\underline{x} = (x_1, \cdots x_n, \ldots, x_N)$), then the bound can be further simplified. Let $P(\underline{x})$ be a probability measure that chooses each letter independently with the same probability (that is, $P(\underline{x}) = \prod_{n=1}^{N} P(x_n)$). Then

$$\overline{P}_e \leq \sum_{\underline{y}} \left[\sum_{\underline{x}} P(\underline{x}) \prod_{n=1}^{N} \Pr(y_n|x_n)^{1/(1+\rho)} \right]^{1+\rho} (M-1)^{\rho}. \tag{13}$$

The term in brackets in Eq. 13 is the average of a product of independent random variables, and is equal to the product of the averages. Thus

$$\overline{P}_e \leq \sum_{\underline{y}} \prod_{n=1}^{N} \left[\sum_{k=1}^{K} P(x_k) \Pr(y_n|x_k)^{1/(1+\rho)} \right]^{1+\rho} (M-1)^{\rho}, \tag{14}$$

where $x_1, \ldots, x_K$ are the letters in the channel-input alphabet. Applying an almost identical argument to the sum on $\underline{y}$, we obtain

$$\overline{P}_e \leq \left[\sum_{j=1}^{J} \left[\sum_{k=1}^{K} P(x_k) \Pr(y_j|x_k)^{1/(1+\rho)} \right]^{1+\rho} \right]^{N} (M-1)^{\rho}, \tag{15}$$

where $y_1, \ldots y_J$ are the letters in the channel-output alphabet.

If the rate is defined in natural units as $R = \frac{\ln M}{N}$, Eq. 15 can be rewritten

$$\overline{P}_e \leq e^{-NE(\rho)} \qquad \text{for any } \rho,\ 0 \leq \rho \leq 1 \tag{16}$$

$$E(\rho) = E_o(\rho) - \rho R \tag{17}$$

$$E_o(\rho) = \ln \sum_{j} \left[\sum_{k} P(x_k) \Pr(y_j|x_k)^{1/(1+\rho)} \right]^{1+\rho}. \tag{18}$$

It can be shown by straightforward but tedious differentiation that $E_o(\rho)$ has a positive first derivative and negative second derivative with respect to ρ for a fixed $P(x_k)$. Optimizing over ρ, we get the following parametric form for $E(\rho)$ as a function of R:

$$E(\rho) = E_o(\rho) - \rho E_o'(\rho) \tag{19}$$

$$R(\rho) = E_o'(\rho) \qquad \text{for } E_o'(1) \leq R \leq E_o'(0) \tag{20}$$

$$E = E_o(1) - R \qquad \text{for } R \leq E_o'(1). \tag{21}$$

From the properties of $E_o(\rho)$, it follows immediately that the E,R curve for a given choice of $P(x_k)$ appears as shown in Fig. X-8. $E_o'(0)$ turns out to be the mutual information on the channel for the input distribution $P(x_k)$. Choosing $P(x_k)$ to achieve channel

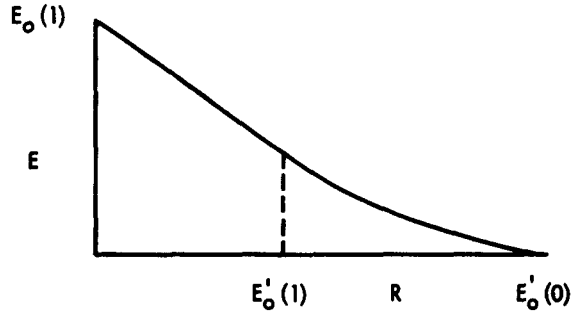


Fig. X-8. Exponent as a function of rate for fixed $P(x_k)$.

capacity, we see that for $R < C$, the probability of error can be made to decrease exponentially with the block length, N. For rates other than channel capacity, $P(x_k)$ must often be varied away from the capacity input distribution to achieve the best exponent.

Equations 19-21 are equivalent to those derived by Fano,[1] and for $E_o'(1) \leq R \leq E_o'(0)$, a lower bound to P_e can also be derived[1] of the form $P_e \geq K(N)\, e^{-NE(\rho)}$, where $E(\rho)$ is the same as in Eq. 19, and K(N) varies as a small negative power of N.

R. G. Gallager

References

1. R. M. Fano, The Transmission of Information (The M.I.T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961), Chapter 9.

XI. ARTIFICIAL INTELLIGENCE*

Prof. M. L. Minsky	D. G. Bobrow	D. C. Luckham
Prof. C. E. Shannon	D. J. Edwards	L. G. Roberts
P. W. Abrahams	T. P. Hart	I. E. Sutherland

A. COMPUTER FORMULATION AND SOLUTION OF ALGEBRA PROBLEMS GIVEN IN A RESTRICTED ENGLISH

A hand analysis of eighty algebra word problems has been made. Several preliminary processing programs were written in COMIT and debugged. It was then found that a more general list-processing language such as LISP would be much more convenient. However, a combination of the two languages would be better than either alone. Therefore, a COMIT interpreter (METEOR) was written in LISP, and debugged. It will handle simple COMIT rules and program flow, with added provision for mnemonic labelling of constituents. Any LISP function can be applied to a constituent, or set of constituents, from the work space. The input-output features of COMIT have been implemented, but the rules are entered in a LISP format. Certain other changes have been made to COMIT format which seemed natural and facilitated programming.

A memorandum has been published by the Artificial Intelligence Group which describes in detail the METEOR program and its use in the LISP system.

D. G. Bobrow

*This work is supported in part by the National Science Foundation (Grant G-16526); in part by the National Institutes of Health (Grant MH-04737-02); and in part by the Computation Center, M. I. T.

XII. SPEECH COMMUNICATION*

Prof. K. N. Stevens	Dr. A. S. House	A. P. Paul
Prof. M. Halle	Jane B. Arnold	T. E. Taylor
Prof. J. B. Dennis	W. L. Henke	S. S. Reisman
Prof. J. M. Heinz		E. C. Whitman

A. ANALYSIS OF ERROR SCORES DERIVED FROM AN AUTOMATIC VOWEL REDUCTION SCHEME

In an evaluation of the performance of an automatic procedure for reducing vowel spectra to descriptions in terms of the time-variant vocal-tract resonances and source characteristics, the behavior of the squared-error score has been examined in some detail. This examination provides some insight into the spectrum-matching accuracy of the automatic program and, it is hoped, may lead also to better understanding of the operational meaning of an adequate match in terms of error score.

The automatic program under investigation has already been described,[1,2] and its accuracy in matching spectra has been discussed[3,4] in terms of differences in pole specifications between vowel samples analyzed by the automatic procedures and by manual procedures. The manual reduction methods are described in detail in a paper by Bell et al.[5] The speech materials are those reported by House,[6] consisting of eight vowels in 14 consonantal contexts spoken by three talkers.

Three contiguous spectral samples located medially in each of the vowels were analyzed manually and automatically, and for each such analysis there corresponds a squared-error score. The mean-error scores associated with each vowel and talker are displayed in Table XII-1 according to spectrum-matching method. Inspection of these scores indicates that the magnitude of the mean error associated with a given vowel, for example, is reasonably stable from talker to talker, and that the average error score of each talker is highly similar. There is, however, a statistically significant difference between the magnitude of the mean-error scores of the two methods, manual and automatic, with the latter providing lower scores. This tendency is quite consistent when scores for individual vowels and talkers are studied. For both the manual and automatic analyses, the vowel means display significant differences, with higher scores tending to be associated with the back, rounded vowels. This tendency may be aggravated when rounded vowels are in an environment provided by labial consonants. In general, however, consonantal context does not provide a significant source of error-score variation.

The mean-error scores per filter for the two methods of analysis are shown in Table XII-2. This presentation emphasizes the consistency with which the automatic

*This research was supported in part by the U. S. Air Force (Electronic Systems Division) under Contract AF 19(604)-6102; in part by the National Science Foundation (Grant G-16526).

Table XII-1. Mean-error scores (in db^2) for spectrum matches to various vowels, arranged by methods and talkers. Scores are computed over filters 1-24 corresponding to a frequency range of 100-3050 cps. Each entry is based on three centrally located samples taken from vowels in 14 consonantal contexts.

Method	Talkers	Vowels								Mean
		i	I	ɛ	æ	ɑ	ʌ	ʊ	u	
Automatic	1	45	34	29	33	30	37	66	55	41
	2	65	42	41	36	46	43	65	56	49
	3	40	35	31	28	34	38	59	65	41
	Mean	50	37	33	32	37	39	63	59	44
Manual	1	49	48	48	48	40	47	80	62	53
	2	49	54	54	46	42	48	53	55	50
	3	44	44	47	39	46	46	59	66	49
	Mean	47	48	49	45	43	47	64	61	51

Table XII-2. Mean error scores (in db^2) per filter for spectrum matches to various vowels, arranged by methods. Scores are computed over filters 1-24 corresponding to a frequency range of 100-3050 cps. Each entry is based on three centrally located samples taken from vowels in 14 consonantal contexts as spoken by three talkers.

Method	Vowels								Mean
	i	I	ɛ	æ	ɑ	ʌ	ʊ	u	
Automatic	2.1	1.6	1.4	1.4	1.5	1.6	2.6	2.5	1.8
Manual	2.0	2.0	2.1	1.9	1.9	2.0	2.7	2.5	2.1

procedure provides lower scores. According to this table the fit of the vowel data is, on the average, accurate to better than 2 db in each of 24 frequency bands.

A. S. House, A. P. Paul

References

1. A. P. Paul, Strategy for the Automatic Matching of Speech Spectra, S. M. Thesis, Department of Electrical Engineering, M. I. T., August 1961.

2. A. P. Paul, Automatic matching of vowel spectra, Quarterly Progress Report No. 63, Research Laboratory of Electronics, M. I. T., October 15, 1961, p. 137.

3. A. P. Paul, An automatic spectrum matching program, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M. I. T., July 15, 1962, pp. 275-278.

4. A. S. House, A. P. Paul, K. N. Stevens, and Jane B. Arnold, Acoustical description of syllabic nuclei: Data derived by automatic analysis procedures, Proc. Fourth International Congress on Acoustics, Copenhagen, August 21-28, 1962 (Paper G22).

5. C. G. Bell, H. Fujisaki, J. M. Heinz, K. N. Stevens, and A. S. House, Reduction of speech spectra by analysis-by-synthesis techniques, J. Acoust. Soc. Am. 33, 1725-1736 (1961).

6. A. S. House, On vowel duration in English, J. Acoust. Soc. Am. 33, 1174-1178 (1961).

XIII. MECHANICAL TRANSLATION*

V. H. Yngve
Carol M. Bosche
Elinor K. Charney
Ann Congleton
J. L. Darlington
Muriel S. Kannel
E. S. Klima
E. S. Lowry
J. D. McCawley
T. More, Jr.
W. K. Percival
A. C. Satterthwait
K. G. Sellin
J. S. Wright

A. TRANSLATING ORDINARY LANGUAGE INTO FUNCTIONAL LOGIC

Work continues on the COMIT program, which was described in Quarterly Progress Report No. 68 (pages 174-175), for translating ordinary language into symbolic logic. The last report described a program for translating sentences and arguments formulated in ordinary English into the notation of propositional logic. Since that time, the author has been working on the extension of that program to perform translations into a "higher" type of logic, i.e., first-order functional calculus. The work that he has done thus far on this subject may be described under three headings: (I) the formulation of a logical grammar, which states the essential logical characteristics of some types of statements in ordinary language which are most readily analyzable into formulae in functional calculus; (II) the writing of a COMIT program that employs this grammar for random generation of sentences; and (III) the writing of a COMIT program for recognition of the grammatical structures of the sentences of the relevant type.

As for (I), the sentences that lie within the scope of our grammar are of two types. In the first type of sentence, an 'NP' (the commonly accepted abbreviation of 'noun phrase') is connected to another NP by a form of the verb 'to be'. In the second type, an NP is linked to another NP by what we may rather sententiously call a "binary relational construction," by which we mean a word or a phrase such as 'helps', 'belongs to', or 'is greater than'. These words and phrases do not fit snugly into any one grammatical category (for instance, they cannot all be described as transitive verbs), but all do the job of stating a comparison or relation between two essentially different NPs. These two sentential types are succinctly stated in two COMIT rules:

(1) S = NP + BE + NP; (2) S = NP + P/.2 + NP

in which 'S' stands for 'sentence', 'BE' for the verb 'to be', 'P/.2' for any binary relational construction, and 'NP', as stated above, for 'noun phrase'. Moreover, 'NP' may be expanded in either of two ways, formulable in the COMIT rules:

(3) NP = CONST; (4) NP = Q/X + PHI/X.

According to (3), 'NP' may be replaced by a proper name, such as 'George Washington'

*This work was supported in part by the National Science Foundation (Grant G-24047).

or 'Chicago'. According to (4), 'NP' may be replaced by a function 'PHI' of X preceded by a determiner or quantifier 'Q' for X. The 'Qs' of X are phrases like 'the X', 'some X', 'exactly five X', and 'at least two X'. The 'PHIs' of X are expressible in the following COMIT rules (a)-(g), in which the only symbol not previously explained is 'P/.1', which stands for nontransitive (or "unary") predicate constructions such as 'is a horse', 'is a Professor of Greek', 'lives', or almost any verb or verbal construction that is not followed by a direct object. (Many verbs, of course, are not definitely classifiable either as transitive or intransitive, e.g., 'believes' and 'sees'. In such cases the recognition program has to look at what immediately follows the verb before opting definitely either for P/.1 or P/.2).

(a) PHI/X = X + P/.1

(b) PHI/X = X + P/.2 + NP

(c) PHI/X = NP + P/.2 + X

(d) PHI/X = X + P/.2 + Y

(e) PHI/X = Y + P/.2 + X

(f) PHI/X = PHI/X + AND + PHI/X

(g) PHI/X = PHI/X + OR + PHI/X

In the formulae above, X and Y may be replaced by any letters from A to Z, subject to certain conditions. Whenever 'NP' is expanded into 'Q/X + PHI/X', in accordance with formula (4), the 'X' chosen is the next letter in alphabetical order; i.e., the first substitution for 'X' is 'A'; the second, 'B'; and so on. (An exception to this rule is made in sentences of type I, in which one may begin expanding both 'NPs' with the same letter 'A'.) The 'Y', in (d) and (e) above, is any letter that has been previously used. The following examples (i)-(vii) are all examples of 'NPs', according to our grammar. They illustrate, respectively, rules (a)-(g) above for the expansion of 'PHI/X'. In each case, the phrase 'such that', although logically redundant, is inserted between 'Q/X' and 'PHI/X' to improve legibility. The generation program, also, employs this device.

(i) The A such that A is a horse.

(ii) At least one D such that D admires George Washington.

(iii) Some B such that Chicago likes B.

(iv) Exactly three C such that C notices A.

(v) All B such that A is greater than B.

(vi) No A such that A is a horse and A is Chicago.

(vii) At most five C such that C is a Professor of Greek or C is a Professor of Latin.

An example of a complete sentence which may be translated into the terminology above is the following, taken from I. M. Copi's Symbolic Logic.[1]

(viii) The architect who designed Tappan Hall designs only office buildings.

Stated in our terminology, it becomes

(ix) The A such that A is an architect and A designs Tappan Hall designs only B such that B is an office building.

It may be noted that two different variables, 'A' and 'B', are employed, and that the imperfect 'designed' is replaced by the present 'designs'. This replacement simplifies the sentence by reducing what are, on the face of it, two transitive verbs to one; and it does not affect the logic of the sentence. An example of a sentence that the machine randomly generated using our grammar is

(x) SIWASH + LIKES + EXACTLY + ONE + X/A + SUCH + THAT + X/A + IS + A + WOMAN + AND + SOME + X/B + SUCH + THAT + X/B + CAN + SWIM + FASTER + THAN + ANY + X/C + SUCH + THAT + X/C + IS + JACK + OF + X/A + CAN + COMMAND + X/A + AND + X/A + GETS + RUSSIA + .

The last example, as opposed to the immediately preceding one, is stated in COMIT notation. We have found it expedient to use as variables 'X/A', 'X/B', 'X/C', etc., rather than 'A', 'B', 'C', etc. In addition to strictly programmatic advantages, this device avoids any confusion between the article 'A' and the letter 'A'.

Next, the recognition program takes sentences such as (ix) and (x) and "parses" them, i.e., it tags each grammatical construction with its appropriate label. It thereby reveals how the sentences must have been, or could have been, generated. Using the recognition program, the computer correctly printed out the analyses of (ix) and (x); these are included below as (xi) and (xii), respectively.

(xi) S + NP + Q/A + THE + X/A + PHI/A + PHI/A + X/A + P/.1 + IS + AN + ARCHITECT + AND/A + PHI/A + X/A + P/.2 + DESIGNS + NP + TAPPAN/CONST + HALL/CONST + P/.2 + DESIGNS + NP + Q/B + ONLY + X/B + PHI/B + X/B + P/.1 + IS + AN + OFFICE + BUILDING + .

(xii) S + NP + SIWASH/CONST + P/.2 + LIKES + NP + Q/A + EXACTLY + ONE + X/A + PHI/A + PHI/A + PHI/A + X/A + P/.1 + IS + A + WOMAN + AND/A + PHI/A + NP + Q/B + SOME + X/B + PHI/B + X/B + P/.2 + CAN + SWIM + FASTER + THAN + NP + Q/C + ANY + X/C + PHI/C + X/C + P/.2 + IS + JACK + OF + X/A + P/.2 + CAN + COMMAND + X/A + AND/A + PHI/A + X/A + P/.2 + GETS + NP + RUSSIA/CONST + .

It may be noted that the recognition program erases the redundant phrase 'such that' whenever it encounters it in the input sentences. Also, it adds some additional subscripts. Proper names are subscripted with 'CONST' (for 'constant'), and conjunctions ('AND' and 'OR') are subscripted with the letter possessed by the two 'PHIs' which they connect. To discover the scope of any 'S', 'NP', or 'PHI', one may employ the following counting method: Starting immediately to the left of the symbol whose scope is to be

discovered, initialize the count at 1 (in COMIT, one would use a marker with the numerical subscript of 1, e.g., 'M/.1'), and proceeding from left to right raise or lower the count in accordance with the rules:

PHI , S = +2; NP = +1; P/.2 , X , AND , OR , BE = -1;
P/.1 , any consecutive string of items labelled 'CONST' = -2;
all other items = 0.

The scope of the term in question is determined as soon as the count 0 is reached. More precisely, when the count 0 is reached, one continues counting items as long as the count remains 0. When an item is encountered which causes the count to deviate from 0, the counting stops, and the scope is taken to be everything except the last item. In other words, if 'P/.1' causes the count to go to 0, the counting continues until it includes everything labelled 'P/.1', such as 'IS + A + HORSE'.

The counting method just described for determining the scope of a term has not yet been programmed, but it will be an essential part of the next program that we intend to write, which will translate the parsed sentences, such as (xi) and (xii), into purely logical notation. The next task following the completion of this will be to write a program that will translate sentences stated in the ordinary English of (viii) into the quasi-logical language of (ix) and (x). These two uncompleted programs, in conjunction with those already written, will constitute a complete program for translating our particular types of sentences from ordinary English into first-order functional calculus.

J. L. Darlington

References

1. I. M. Copi, Symbolic Logic (Macmillan Company, New York, 1958), p. 158.

B. FUNDAMENTAL SENTENCE-MEANING

In previous Quarterly Progress Reports[1-3] I have pointed out that a general explanation of the phenomenon of linguistic paraphrase cannot be achieved within the framework of a theory of semantics that makes the tacit assumption that word-synonymy (or morpheme-synonymy) alone can account for it. Such a view not only leads inexorably to the absurd conclusion that a structural-constant, such as 'all', 'only', 'any', 'some', 'not', 'ever', must at times be considered as synonymous with one structural-constant and at other times synonymous with a different structural-constant, but fails to provide the technical means powerful enough to discover fundamental semantical relations that exist among these structural morphemes, relations that can be formulated into linguistic laws that are general in that they encompass vast areas of linguistic phenomena,

and are explanatory in that they provide reasons why one cannot always substitute the same structural-constant for another in all sentences containing an occurrence of that morpheme and obtain sentences that are synonymous, why and when there occurs a change of meaning if one makes morpheme substitution of this kind, and so on. The reason why a method of discovery based upon such a view of semantics cannot be powerful enough to discover more fundamental laws is that it has been, by definition, restricted to the use of only substitution rules for the greater part. That is to say, on the basis of this assumption concerning semantics, the structure of the grammatical strings – the very structure that is to be investigated – has been predetermined to be essentially of the phrase-structure type, wholly accessible to constituent analysis, with some rules of ordering of constituents thrown in when necessary. The type of grammar has already been assumed to be a context-free grammar.

In the theory of semantics proposed by the author, it is held that the word-meaning of these critical structural morphemes always remains constant – hence the name structural-constant – and that the sentence-meaning can vary according to the particular configuration of structural-constants which occurs in the sentence. Some configurations give rise to sentences whose sentence-meanings are synonymous although there is no one-to-one morpheme correspondence, and others give rise to sentences whose sentence-meanings are different. In order to illuminate how the meaning of a semantical structural-constant can remain constant and at the same time the sentence-meaning can vary depending upon the structural context, let us construct the following analogy.

Let us liken the structural configuration of a sentence-type to a machine constructed out of cogwheels, levers, rods, bolts, screws, and other such parts. The particular set of parts and the manner in which they are connected to one another give rise to a machine that has a certain output. The same parts, when combined in a different way, give rise to a different machine. It may happen that particular choices of parts and different combinations of them give rise to several machines that operate to produce the same output, although they are not structurally isomorphic; these could be called synonymous machines. In this case it could be said that the machines have the same fundamental construction. In the case in which several machines are not structurally isomorphic and do not operate to produce the same output, the machines are said to have different fundamental constructions. A construction engineer knows which choice of parts and which combinations of them give rise to the various kinds of machines. In the case in which the engineer has built two or more machines that appear to be structurally isomorphic but operate to produce different output, some of the machine parts have become built-in units; the structures that differ are hidden under the cover of a closed unit. Such machines could be said to be ambiguous.

It is obvious from the analogy that the author believes that only a context-sensitive grammar is adequate to handle the structure of language. The analogy is so obvious

that one need not dwell on it. But I would like to discuss briefly the semantical structural-constants, 'all' and 'only'. It is only when one regards 'all' as a constant, analogous to a functioning meaningful machinery part, that one can explain why 'all' can range in meaning over 'everything', 'the totality', 'the whole', 'the end', etc. As an operator whose function is to determine the extent of the upper boundaries enclosing all members from the inside to the outer limits, it can range over many different types of things; sometimes it operates on countable argument individuals, sometimes on predicate functions such as verbs, adjectives, and nouns, sometimes on functions of functions such as adverbs. Sometimes it operates upon the infinitely divisible parts belonging to a whole or single individual. But it always has one job to do, a job that never varies. The morpheme 'only' also functions to determine a boundary, but, unlike 'all', it does not specify that the total membership is involved; rather it restricts, or clamps the lid down on the upper boundary from outside the set. 'All' and 'only' are closely related operators in that they operate in the reverse of each other. Examples (1) and (3) have been selected to illustrate the comparison. Since, however, there does not exist a logical operator corresponding notationally to the grammatical form of 'only', 'only' is expressed symbolically in the predicate calculus through a paraphrase of the sentence containing an occurrence of 'only', a paraphrase in which 'all' appears. The English language often makes use of this paraphrase to express 'only'. Illustrations of this usage are given in examples (7) and (9) and again in examples (20a) and (20b). In order that people untrained in the logical symbolism of the predicate calculus can follow the text more easily, I shall not use symbolic notation in this report. However, it must be kept in mind that for each sentence in the examples there is a general schematization, a symbolic formula in which the denotative terms of the sentence are expressed as variables, only the range of the variable having structural significance; this formula represents the sentence-type of the sentence in question.

In order to see the effect that 'only' can make on the fundamental sentence-meaning, let us first examine an example of a rather simple sentence-type whose subject noun phrase contains a universal quantifier,

(1) All women are frail

whose fundamental sentence-meaning is

(2) To be a woman is a sufficient condition of being frail.

If we replace 'all' by 'only' in sentence (1), we obtain

(3) Only women are frail

whose fundamental sentence-meaning is the reverse of sentence (2), namely,

(4) To be a woman is a necessary condition of being frail.

The morpheme 'only' functions here to reverse the order of the asymmetrical operation of implication. In other words, if 'all' is the major quantifier of a sentence S that can be represented by the general formula in the predicate calculus '$(x)[f(x) \supset g(x)]$' (read: for every x, if x is f then x is g), then if 'only' is the major quantifier of a sentence S' containing all the same morphemes as S with the exception of 'all', then the general formula is written '$(x)[g(x) \supset f(x)]$' (read: for every x, if x is g then x is f). Inspection of the symbolic forms of these two formulas reveals the reversal in the order of the left-hand term before the symbol for implication and the right-hand term after the symbol.

Given an example of a sentence-type like

(5) John needs to move to feel pain

which already expresses necessary condition, since its fundamental sentence-meaning is

(6) Moving is a necessary condition that John must satisfy in order to feel pain

we see that the addition of 'only' after 'needs' in sentence (5) serves again to reverse the fundamental sentence-meaning. The resulting sentence

(7) John needs only to move to feel pain

expresses the fundamental sentence-meaning

(8) Moving is the only condition that John must satisfy in order to have pain.

The reversal that the addition of 'only' accomplishes is seen in the fact that sentence (8) implies that not only is moving a necessary condition of feeling pain but a sufficient condition; that is, since moving is the only condition needed, it is the necessary and sufficient condition of producing pain. Sentence (7) is synonymous with

(9) All John needs to do is to move to feel pain

since both express the same fundamental sentence-meaning. The fundamental sentence-meaning could be expressed in a grammatical form notationally closer to sentence (9) by

(10) All conditions that John must satisfy in order to feel pain are the same as to move.

Incidentally, note that both sentence (7) and sentence (9) are synonymous with

(11) John need only move to feel pain

and

(12) All John need do to feel pain is to move

in both of which 'need' is clearly a modal. Compare sentences (11) and (12) with the sentence-type of sentence (5) in which there is not the option of using the modal form or the verbal form of 'need'; the verbal form is obligatory.

By understanding the role that the semantic and purely syntactic structural-constants play in giving rise to sentence-meaning, there does not seem to be anything more mysterious in the concept of sentence-synonymy than there is in two maps, very different in outward shape, representing the same country through different projections. As long as there exists a one-to-one mapping between the projections, the two maps are regarded as fundamentally isomorphic. In the investigation of fundamental semantical relations of a natural language system, we search for translation rules that transform one sentence into another sentence that is synonymous with it. These translation rules are the mapping rules of the linguist interested in a semantically grounded grammar.

Every translation rule is justified as a correct rule if and only if there corresponds to it a metalanguage statement, empirically established as true, which states the fundamental sentence-meaning that the synonymous sentences have in common. Basic translation rules are, of course, not easy to discover, particularly if the problem is attacked by a hit-or-miss, trial-and-error method. Since the domain of linguistic data is so immense, a procedure without a logically directed program of attack can be quite fruitless. Fortunately, one can utilize the knowledge gained from the advances made in the understanding of the logical structure of artificial language systems such as the predicate-calculus to construct such a program, even though one knows from experience that a natural language system is not as limited in scope or purpose as an artificial one. If we regard the artificial logical calculi as analogs of natural language systems, we are justified in assuming that there must exist a metalogic for any language system, that is, a body of implicit basic rules governing the use of the conventional symbols in the language: implicit derivation rules, formation rules, and axioms. If there did not exist such a logical set of rules governing the grammatical forms of sentences, it would be difficult to understand how a language could be mastered, since it would then not be a system but a very large set of unconnected strings. When one masters a language one has learned how to use these rules; but there is a difference between being able to use a rule and being able to state the rule explicitly. Observation and intuitive knowledge of the language in question can lead to the establishing of a set of explicit translation rules, accepted as correct because many sentences have been tested empirically to confirm the fact that the structural-constants do, in fact, behave in the way stipulated. Thus when one has stated some of the metalogical rules, one can use them to demonstrate, according to the metalogic of that language, that certain sentences are synonymous. That is, one can validate the intuitive recognition of synonymy of several sentences by logical proof, a logical proof that utilizes only a partial knowledge of the meanings of some important structural-constants in that language, the meanings of the denotative constants, and the

fundamental meanings of a few well-chosen sets of synonymous sentences regarded as axioms, and the logical rules of deductive inference. If the synonymy of sentences can be validated, one is confirmed in believing that the rules already formulated are actually basic to that language system. Furthermore, new structural-constants and new rules can be discovered in terms of the old. The following examples have been selected to illustrate how English sentences, found intuitively to be synonymous, can be proved to be synonymous by a tentatively accepted metalogic of English.

The concepts of upper and lower bounds play an important role in the English language. Since the quantifier 'all' is essential in symbolic formulas expressing ideas concerning limits, let us compare some sentences containing different grammatical occurrences of 'all' with their respective intuitively synonymous sentences.

(13a) I helped him all I could
(13b) I helped him as much as I could

(14a) St. Paul is the farthest I will go
(14b) I will go only as far as St. Paul
(14c) St. Paul is all the farther I will go

(15a) What big eyes you have! All the better to see you, my dear
(15b) What big eyes you have! So much the better to see you, my dear

(16a) He was all but smiling
(16b) He was almost (nearly) smiling

(17a) He was anything but smiling
(17b) He was far from smiling

(18a) All I could do was to smile
(18b) I could only smile

(19a) It was all I could do to smile
(19b) I could hardly smile

(20a) I was all afire
(20b) I was completely afire

(21a) All of the class was excited
(21b) The whole class was excited

According to grammarians such as Poutsma, 'all' in sentence (13a) is to be equated with 'as much as' in sentence (13b), since these words can be interchanged with the meaning of the original sentence preserved. According to the principle of sentence-synonymy, the two sentences as wholes have the same fundamental sentence-meaning but no synonymy holds between 'all' and 'as much as'. In sentence (13a), the universal quantifier ranges over the set of possible degrees of help that x_1 was able to give; thus x_1 helped x_2 with every degree of help that it was possible for x_1 to give x_2. If x_1 helped with all possible degrees, x_1 helped with the highest degree as well. That is to say, x_1 helped up to the limit of his ability. Another sentence that is synonymous with

sentences (13a) and (13b), which corresponds to this last statement that was derived by logical inference, is 'I helped him the most I could'. Sentence (13b), on the other hand, states that the amount of help given to x_2 was 'equal to' or 'as much as' the amount x_1 was able to give. Since the amount that it was possible for x_1 to give is, in fact, the greatest amount, more being impossible, the amount of help given equals the largest or the upper limit. Ergo, the fundamental sentence-meaning for all three statements has been proved to be identical. Their sentence-types, however, are different. It is to be noted that the verb 'help', conceived as having different intensities ordered as to degree, is a necessary assumption to the explanation of the sentence-synonymy. The English language makes much use of the notion of ordering, as we shall see in examples (14)-(19).

The sentence-meaning of sentence (14a) is a very direct expression of the fundamental sentence-meaning of all three examples in set (14), which is

(22) St. Paul is the upper limit of the distance from an initial point that x_1 is willing to go

because the sentence itself identifies by means of the word 'is' the place, St. Paul, with 'the point that is farthest', the use of the superlative guaranteeing the uniqueness of the point that is farther than all other points x_1 is willing to go to from a given starting point. The sentence-meaning of (14b) states that x_1 is willing to go (from some given point) as far as St. Paul but no farther; thus sentence (14b) states that x_1 is willing to go from an initial point to a point that is 'equal to' or 'as far as' the upper bound, St. Paul, and does not exceed that point. Whether St. Paul is said to be the limit itself, or the distance that x_1 is willing to go is declared to be as far as but not farther than the limit, the distance in both cases is the same; hence the two sentences are synonymous. The grammatical differences and similarities between the two sentences are easily seen when both are put into their schematic formulations.

The proof of the synonymy of sentence (14c) to (14a) and (14b) is more complicated. The phrase 'all the farther' refers to the total increase in distance. The complete meaning of 'x is farther than y' has to be expressed 'x is farther from an initial point y_o than y is by a specific but unknown amount α', i.e., x increases in distance away from y_o more than y increases in distance by a definite increase. It should be noted that in any ordering relation applied to things, whether the ordering be in terms of intensity, degree, or size, there is always an amount involved. Sentence (14c) identifies two classes by means of the word 'is', which stands between the left-hand phrase that defines the set of points, '(from y_o to) St. Paul', and the right-hand phrase that defines the well-ordered set of points, 'all the increases in distance x_1 is willing to go'. The right-hand phrase can be represented as the union of the collection of sets containing as members all of the α's, which are obtained by taking in their ordered

turn all those points that are distant from y_o to which x_1 is willing to go. Thus each new point that x_1 is willing to go to starting from y_o farther than the previous point becomes an upper bound in its turn so that the smallest that each increase can be is simply from one point to the next farther point. The sum of all of these increases (or all farther points) is, of course, equal to the distance from y_o to St. Paul because the sentence has equated these two sets. The sets are identical because each set has exactly the same members. The sentence-type of (14c) is quite different from those of (14a) and (14b). Both (14a) and (14b) correspond notationally to the standard symbolic definition of an upper bound; sentence (14c), by declaring the two sets to be identical, lays down a condition under which there must exist an upper bound to the set of points to which x_1 is willing to go.

In (15a) we have the familiar lines in the nursery tale of Red Riding Hood and the wolf. There is again ordering according to size; but in this example two properties are being ordered and the ordering is correlated, since the value of the increase or decrease in one variable depends upon the determination of the increase or decrease in the other. The wolf's assumption is

(23) The bigger the eye, the better the sight of the eye.

The fundamental sentence-meaning of (23) is that the goodness of the sight of the eye increases as a function of the greatness of size of the eye. In sentence (23), the independent variable is given in the first phrase, the dependent in the second, so that the ordering of the two phrases is an indispensable feature of the sentence-meaning. If the order is reversed, grammatical changes occur; the synonymous sentence would then be

(24) The sight of the eye is better, the bigger the eye.

When Red Riding Hood asserts that Grandma's eyes are large, she means that as eyes they have the property of largeness, since size is never an absolute property but is relative to the kind of object talked about. Hence the word 'large' must be defined 'large with respect to some standard'. Grandma's eyes being a specific amount, although an unknown amount, larger than average, the ability of her sight is directly correlated to that amount and can, in theory, be calculated. The phrase 'all the better to see you' in sentence (15a) is similar to the phrase 'all the farther' in the previous example, since it too denotes the set of 'all the increases in goodness of seeing'. The sum of all these increases equals the amount that her eyes are larger than average. Hence sentence (15b) that asserts 'that much (or so much) better to see you' is synonymous with sentence (15a) because the union of the set of all the increases equals that definite amount, the difference in size of eye; hence 'so much the better' can replace 'all the better'.

Examples (16)-(19) illustrate that the English language possesses techniques for handling the concept of proximity to total absence or total presence of a property. Does a person either smile or not smile? Our language is able to construct statements expressing how close to the boundary of yes or no a fact may be. In sentence (16a) x_1 is not smiling but very near to the upper bound of the separation between not smiling and smiling. The definition of the complex quantifier 'almost' involves the occurrence of the universal quantifier 'all' in combination with the structural-constant 'except' or 'but' to exclude a limited finite number or amount from the total set – an amount that approaches but never reaches zero as the limit. Hence, the words 'all but' correspond directly to the definition of 'almost'.

Sentence (17a) states that x_1 was not smiling but states that every property that could be included as part of smiling was absent; hence x_1 is far from the upper bound of the separation between smiling and not smiling.

The synonymy of sentences (18a) and (18b) was treated in detail in Quarterly Progress Report No. 66 (pages 289-293). I included them in this report to complete the list of different grammatical occurrences of 'all' and to bring into sharp focus the critical effect resulting from additional words such as 'It was' placed in front of sentence (18a), as in example (19a). Sentence (19a) expresses the sentence-meaning that although x_1 smiled, it was necessary for him to use the limit of his ability to do so. Thus x_1 was close to the lower limit of smiling, that is, close to not smiling.

The synonymy of sentences (20a) and (20b) is demonstrated by showing that the universal-quantifier 'all' ranges over all of the parts of an individual. Since 'all' is a unity, it includes all of its parts and is therefore complete.

The explanation of the synonymy of sentences (21a) and (21b) is similar to the explanation of set (20). It should be pointed out that sentences (20a) and (20b) are synonymous with 'I was wholly afire'. In set (21), the universal-quantifier 'all' again operates upon the parts of an individual. In the case of sentence (21a) the use of the definite article before 'class', a collective noun, shows that 'class' is regarded as singular. 'All of the class' and 'the whole class' refer to the class as a complete unity, that is, the individual composed of all of its parts.

It has perhaps been remarked that the phrase 'fundamental sentence-meaning' has been used frequently. What is the exact meaning of this phrase and just what is the relation between fundamental sentence-meaning and ordinary sentence-meaning? In this report I would like to discuss and clarify this elusive concept and try to explain its role in relating sentences as synonymous.

In Quarterly Progress Report No. 68 (pages 176-180) I stated that the fundamental sentence-meaning was logical in form, not grammatical; it was the forms of the sentences found to be synonymous by empirical observation which were grammatical. On the basis of my present point of view, I would not now deny grammatical form to the

sentence stating the fundamental sentence-meaning of a set of sentences that are alike in expressing that sentence-meaning. Rather, I would describe the relationship of this sentence to the others as follows. While it is true that fundamental sentence-meaning is a concept belonging properly to the field of logic, this fact means that the sentence stating the fundamental meaning of a set of synonymous sentences belongs properly to the metalanguage of our language system; the synonymous sentences referred to by the metalinguistic statement belong to the object language of English. The metalinguistic statement states explicitly, by means of denotative terms, the specific kind of relation that exists between the events or things designated by specific types of object-language sentences. That is, the metalinguistic sentence, by stating which basic relation the object-language sentences express, is using the name of this particular relation denotatively; thus, for example, the phrase 'necessary condition' is itself not a structural-constant but a denotative term, defined lexically in the same manner as other nouns, and behaving grammatically as one. Like all other sentences, the metalanguage sentence has a grammatical structure that can be described, and a sentence-meaning of its own. In fact, it is possible for the metalanguage statement itself to be synonymous with other metalanguage statements if the denotative constants are the same, as we saw in examples (8) and (10). On the other hand, the object-language sentences that the metalanguage sentences refer to as being synonymous do not use the denotative term to refer reflexively to themselves; they express the particular relationship that they have in common by using the proper structural-constants in specific configurations. They are sentences in the object language because they use structural-constants that can be defined only in the metalanguage. Let us inspect two synonymous sentences in the object language that are grammatically similar in that all of their denotative terms are the same, such as

(25a) If John is to be president, he must get his organization ready now
(25b) Only if John gets his organization ready now can he be president

and compare their grammatical forms with the grammatical form of the metalanguage sentence (26) stating their common fundamental sentence-meaning

(26) The fundamental sentence-meaning of 'If John is to be president, he must get his organization ready now' and 'Only if John gets his organization ready now can he be president' is that John's getting his organization ready now is a necessary condition for John's being president.

Sentences (25a) and (25b) do not explicitly state that one event is a necessary condition of the other; rather each expresses this fact through its grammatical form. On the other hand, sentence (26), although it too expresses through its grammatical form a fundamental sentence-meaning that could be made explicit by a meta-metalanguage statement, states explicitly which fundamental sentence-meaning examples (25a) and (25b)

have in common, and the fundamental sentence-meaning of metalanguage sentence (26) is not the fundamental sentence-meaning referred to by its subject noun phrase. However, if we look at only the clause 'John's getting his organization ready now is a necessary condition for John to be president', we see that this part of sentence (26) has the grammatical structure of a complete sentence and it is synonymous with sentences (25a) and (25b) even though its denotative terms are not everywhere the same as those of (25a) and (25b). Indeed, if this sentence were not synonymous, it could not have been identified with the fundamental sentence-meaning of these object-language sentences. It appears, then, that one sentence out of a set of synonymous sentences – a wider set that includes synonymous sentences whose denotative terms are not everywhere the same – has been selected as the sentence whose ordinary sentence-meaning is regarded as expressing the fundamental sentence-meaning of the set. That the selection of this one sentence is not arbitrary is seen by the fact that it alone (or synonymous sentences whose sentence-type may differ but whose denotative terms are everywhere the same), because of the type of grammatical structure that it has, can appear intact in a metalanguage statement of the sentence-type represented by sentence (26) in which the subject, which refers denotatively to the concept 'the fundamental sentence-meaning' specifically identifies this sentence with it. It is obvious that neither of the object-language sentences (25a) or (25b) could serve grammatically in this capacity. It is by this grammatical criterion that the sentence-type of a sentence expressing the fundamental sentence-meaning can be objectively established as of a higher structural level than the ordinary object-language sentences.

It should be remarked that the infinite regress that one is forced into by admission of a hierarchy of levels of language is not so ominous as it appears, since the grammatical structure of all levels has been assumed, on the basis of experience with levels already constructed, to be similar.

The fundamental sentence-meaning, I suspect, is what has often been called the 'proposition' that sentences signify or designate. There has been much philosophical discussion concerning the nature and ontological status of the so-called proposition. I would not like to regard the proposition as a logical entity that is incapable of having grammatical form. Such a view would make a mystery of what the proposition is; it seems dubious that it could be logical in character if it cannot be expressed in language.

Once it is recognized that sentences do have a sentence-meaning that is quite distinct from word-meaning, that there is synonymy of sentence, as well as synonymy of word, we no longer need the 'abstract proposition' to relate sentences having the same meaning to one another; the concept of fundamental sentence-meaning, which states what the synonymous sentences express, suffices. (Incidentally, if it is possible that a sentence expresses a sentence-meaning that is not synonymous with that of any other sentence of the same level, its fundamental sentence-meaning must still be stated in the

metalanguage of that language, since the metalanguage statement would utilize denotative terms to refer explicitly to the meaning that the structural configuration of that sentence expresses.)

The concept of 'proposition' referred not only to the entity that synonymous sentences within the same language system designated; it stood also for the entity that related sentences having the same meaning but belonging to different language systems because it was felt that there must be some absolute underlying logical construction that these sentences designated which made possible translation from one language to another. Since the proposition could not be expressed in any of these concrete languages because then it would be a sentence of that language, it was deemed to be not expressible in any natural language; it stood between any two languages. It was supposed that the predicate calculus itself was the underlying logical structure of all natural language systems, that the predicate calculus played a role similar to the mathematical symbolism used universally by mathematicians, each of whom could use the symbolism within his own native language. The inadequacy of the predicate calculus to fill this role has been amply demonstrated.

According to this theory of the semantics of sentence-meaning, the concept of fundamental sentence-meaning is instrumental in character, not absolute, since it is fundamental with respect to some language system, determined implicitly by the formation rules, derivation rules, and axioms of its metalanguage. The applicability of any logical system depends upon how well it serves its recognized purpose, whether it be a calculus constructed by known men or a naturally developed language constructed gradually by many unknown men; therefore it is not surprising that the basic foundations of the highly developed language systems that these philosophers and mathematicians had command of contained concepts mutually recognized as important.

It should be noted that the basic concepts referred to by the fundamental sentence-meanings are man's constructs, not things of the world; they are man's way of looking at the world, organizing its material in ways useful to him. The fundamental sentence-meanings refer to such concepts as necessary and sufficient conditions, causality, generalization, unexpected occurrences, and possibilities. Since these are man's concepts, a pertinent question that one should raise about the possibility of writing contrastive grammars for languages in general and about the possibility of mechanical translation from any language to any language, in particular, is: Are the fundamental sentence-meanings belonging to each of the natural languages that are expressible by the manifold linguistic devices displayed by each of these widely differing grammars, sufficiently universal to serve as the mapping functions with which to coordinate the sentence-types of any language system with those of any other language system? Could there be such wide agreement about what concepts are fundamental even though the cultural environments of man are so diverse? I think one can make an educated guess

based on empirical evidence that it is quite likely that there would be considerable overlap of the fundamental sentence-meanings belonging to different language systems. During the natural weeding-out process that must have taken place during the development of a language, it must have been found out from experience that knowledge of some concepts served man's purposes better than others; hence the general adoption of grammatical techniques to express these important ideas was not just arbitrary; they were selected in terms of their being the best instruments to reach a goal. The goals determining the selection processes are the primary needs and purposes of man. And his most fundamental need is to get along in his environment. Language then has to be an instrument that is capable of communicating knowledge that members of a language community need to express and wish to express to one another. To help one another to adjust to nature, to control it to some extent, to make plans, men need to predict it, to have some sense of order, to construct laws about it, and so forth. Concepts like necessary and sufficient conditions, concepts necessary for the possibility of making inferences, concepts like negation and affirmation that permit the possibility of distinguishing between truth and falsehood would all appear universally needed by a society that is intelligent enough to create and use such an abstract and complicated system of conventional signs as a language. Also, as our biological knowledge of the organization of the human brain increases, it may at some future time be shown that, in the evolutionary process of adaptation, the human brain became structured in such a way that it organizes the raw material of the senses in specific ways.

An interesting project would be a comparison in the sophistication of the concepts involved in fundamental sentence-meaning – all languages possess fundamental sentence-meanings, if not the same ones – between the language of a highly civilized society and the language of what is considered by anthropologists to be a very primitive society.

Elinor K. Charney

References

1. Elinor K. Charney, Linguistic analogues of the free-variable, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M.I.T., January 15, 1962, pp. 208-211.

2. For examples of 'all' and 'only', see Elinor K. Charney, On the problem of sentence synonymy, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 289-293.

3. Elinor K. Charney, Sentence-meaning and word-meaning, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M.I.T., January 15, 1963, pp. 176-180.

XIV. LINGUISTICS*

Prof. R. Jakobson	Dr. K. Wu	S.-Y. Kuroda
Prof. A. N. Chomsky	T. G. Bever	D. T. Langendoen
Prof. M. Halle	S. K. Ghosh	T. M. Lightner
Prof. Y. Isami	Barbara C. Hall	P. M. Postal
Dr. G. H. Matthews	J. J. Katz	C. B. Qualls
Dr. Paula Menyuk	R. P. V. Kiparsky	J. J. Viertel

A. THE NATURE OF A SEMANTIC THEORY

As an investigation in semantic meta-theory,[1] this report describes the abstract form of a semantic theory (or description) of a natural language: the form of the dictionary entries, the rules that project the lexical meanings expressed in dictionary entries onto sentences in the form of sentential meanings, the relation of such rules and dictionary entries, both to each other and to the grammar of the language, and the notion 'sentence meaning'. A meta-theory for semantic theories is needed to inform the field linguist of the types of semantic facts for which to look, the most revealing and succinct way to arrange them, and that which can be said about the semantic structure of the language based upon such facts in such an arrangement.[2]

In our paper "The Structure of a Semantic Theory" (henceforth we shall refer to this publication as "SST"), Fodor and I show that a semantic theory of a natural language has as its fundamental aim the construction of a system of rules which represents that which a fluent speaker knows[3] about the semantic structure of his language which permits him to understand its sentences. The idea behind this conception of a semantic theory is that such knowledge takes the form of recursive rules that enable the speaker to compose, albeit implicitly, the meaning of any sentence out of the familiar meanings of its familiar elementary components. This idea has the following two-part rationale. First, the most salient and impressive fact about linguistic competence is that a fluent speaker can understand a sentence of his language even though he has never previously encountered it. In principle,[4] he can understand any of the infinitely many sentences of his language. But, since at any time in his life the speaker can have encountered only an exceedingly small, finite subset of the infinite set of sentences of his language, and, moreover, his storage capacity is finite, we can conclude that the speaker's knowledge of the semantic structure of his language takes the form of a finite set of recursive rules that fix a meaning for each of the infinitely many sentences of his language. Second, since a speaker's understanding of the sentences of his language also depends on his knowing sufficiently the meanings of their elementary components, the lexical items in the vocabulary of the language, we can conclude that the meaning that the rules fix for a sentence must be a compositional function of the antecedently known meanings of its elementary components, the lexical items appearing

*This work was supported in part by the National Science Foundation (Grant G-16526); in part by the National Institutes of Health (Grant MH-04737-02); and in part by the U. S. Air Force (Electronics Systems Division) under Contract AF19(628)-2487.

in it. Hence, a semantic theory of a natural language must contain rules that explicate the compositional function that determines how a speaker utilizes the meanings of the lexical items in a sentence to understand what that sentence means. If a semantic theory is not adequate to explicate this function, it cannot represent the speaker's knowledge of the semantic structure of his language.

In SST, we proposed that a semantic theory consist of two components. First, a dictionary that provides an entry for each lexical item of the language which, in some sense, gives the meaning of the entry. Second, a finite set of what we called "projection rules" that use lexical information supplied by the dictionary entries for the lexical items in a sentence and information about the sentence's syntactic structure supplied by its grammatical description in order to assign a semantic interpretation to the sentence.

Since information about a sentence's syntactic structure will be needed to assign a semantic interpretation to it, we found it convenient to let the output of a grammar be the input to a semantic theory. In this way, each sentence considered by a semantic theory is represented as a concatenation of morphemes whose constituent structure is given in the form of a hierarchical categorization of the syntactical parts of the concatenation.[5] The sentence The boys like candy is represented by the concatenation of morphemes the+boy+s+like+candy which is hierarchically categorized as follows: the whole string is categorized as a sentence at the highest level of the hierarchy; the+boy+s is categorized as a noun phrase, and like+candy is categorized as a verb phrase at the next level of the hierarchy; the is categorized as an article; boy+s is categorized as a noun; like, as a verb; and candy, as a noun; and so forth on the next and lower levels of the hierarchy. Following Chomsky, we can represent such a categorization in the form of a labelled tree diagram in which the notion 'the sequence of morphemes m belongs to the category c' is formalized by the notion 'm is traceable back to a node labelled c'.[6] Calling this element of the structural description that the grammar assigns to a sentence its "constituent structure characterization," we have the constituent structure characterization of The boys like candy given in Fig. XIV-1.[7] Therefore the input to a

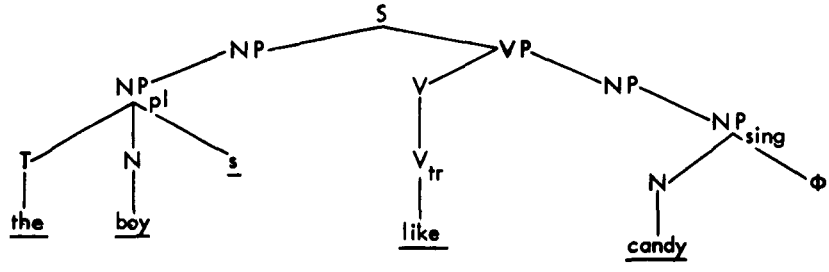


Fig. XIV-1.

semantic theory is sentences thus embedded within their constituent structure characterizations, together with any further grammatical information that an optimal grammar will supply about them.[8]

A semantic theory takes, one after another, the discrete outputs of the grammar and operates on them in a way that matches what a fluent speaker does to obtain his understanding of the sentences. This way of conceiving of the operation of a semantic theory provides a basis for explaining how a fluent speaker applies his stock of lexical information to apprehend the meaning of each new syntactically well-formed arrangement of lexical items that he encounters. Clearly, a semantic theory can provide such an explanation only if its projection rules assign each sentence a semantic interpretation that represents the manner in which a fluent speaker employs the syntactic structure of the sentence to determine its meaning as a function of the meanings of the sentence's lexical items. Hence, we must consider just what a semantic interpretation ought to tell us about a sentence.

The semantic interpretations produced by a semantic theory for the sentences of a language constitute the theory's description of the language's semantic structure. Since a fluent speaker's knowledge of the semantics of his language manifests itself in his verbal performance, a semantic interpretation of a sentence ought to tell us whatever the speaker implicitly knows about the sentence's semantic structure that enables him to carry on his verbal performances. Thus, the fundamental question we asked in SST about the speaker's verbal performance was "What does the speaker do that manifests his knowledge of the semantic structure of his language?" We answered that he differentiates sentences that are semantically acceptable from those that are semantically anomalous, he recognizes ambiguities stemming from semantic relations in a sentence, he detects semantic relations between expressions and sentences of different syntactic type and morphemic constitution, and so forth. On the basis of this answer, we concluded that the semantic interpretations produced by a semantic theory must mark as semantically acceptable and anomalous those sentences that the speaker differentiates as acceptable and anomalous, mark as semantically ambiguous those sentences that the speaker regards as such, mark as semantically related in such-and-such a fashion just those n-tuples of expressions and just those n-tuples of sentences that the speaker detects as so related, and so forth. Otherwise, we argued that the semantic theory cannot claim to represent the speaker's semantic knowledge. For example, a semantic theory of English would have to produce: a semantic interpretation for The bank is the scene of the crime that marks it as semantically ambiguous; semantic interpretations for the sentences He paints with silent paint and Two pints of the academic liquid![9] that mark them as semantically anomalous; semantic interpretations for He paints silently and Two pints of the muddy liquid! that mark them as semantically acceptable; and semantic interpretations that mark the sentences Eye-doctors eye blonds, Oculists eye blonds, Blonds are

eyed by eye-doctors, etc. as paraphrases of each other but mark Eye-doctors eye what gentlemen prefer as not a paraphrase of any of these sentences.

To finish describing the conception of a semantic theory developed in SST, we need only characterize the notions 'semantic interpretation', 'dictionary entry', and 'projection rule'. However, it is important to point out that our characterization of these notions must be such that the semantic interpretations, dictionary entries, and projection rules of a semantic theory of any natural language will be specified formally, i.e., the application of the projection rules will be determinable solely on the basis of the shapes of the symbols in the strings to which they apply and the operations that the rules affect upon these strings will be mechanical. The degree to which a semantic theory is not formally specified is the degree to which the fluent speaker's linguistic knowledge, or his intelligence and ingenuity, are needed to determine whether or not the projection rules apply in certain cases and to determine what operations these rules affect when they apply. Hence, to require that a semantic theory of a natural language be fully formal is just to require that it fully explicate the speaker's knowledge of the semantic structure of the language.

Let us start with the notion 'dictionary entry'. As we have seen above, a semantic theory is intended to reconstruct the process by which a speaker projects a meaning for a sentence from the meanings of the lexical items appearing in it. Thus, within a semantic theory the dictionary entries play the special role of providing the basis from which the projection rules of the theory derive the semantic interpretations that they assign sentences. Our characterization of the notion 'dictionary entry' must thus be such that in it we have a normal form for the dictionary entries which enables us to represent lexical information in a manner that is both formal and sufficient in content to provide a complete basis from which the projection rules can operate.

For the vast majority of cases,[10] a dictionary entry is a set consisting of a finite number of sequences of symbols, each sequence consisting of an initial subsequence of syntactic markers, followed by a subsequence of semantic markers, then, optionally, one distinguisher, and, finally, a selection restriction. Dictionary entries can be represented in the form of tree diagrams, such as that shown in Fig. XIV-2, in which each sequence in the entry for a lexical item m_i appears as a distinct path rooted at the lexical item m_i.[11] As illustrated in Fig. XIV-2, semantic markers are represented enclosed within parentheses, the distinguishers are represented enclosed within brackets, and the selection restrictions are represented within angles. Each complete path, each sequence, represents a distinct sense of the lexical item in whose entry it appears. Thus, in Fig. XIV-2 the lexical item bachelor is represented as having four distinct senses.

Semantic markers are the formal elements that a semantic theory employs to express semantic relations of a general nature. In contrast, distinguishers are the formal elements employed to represent that which is idiosyncratic about the meaning

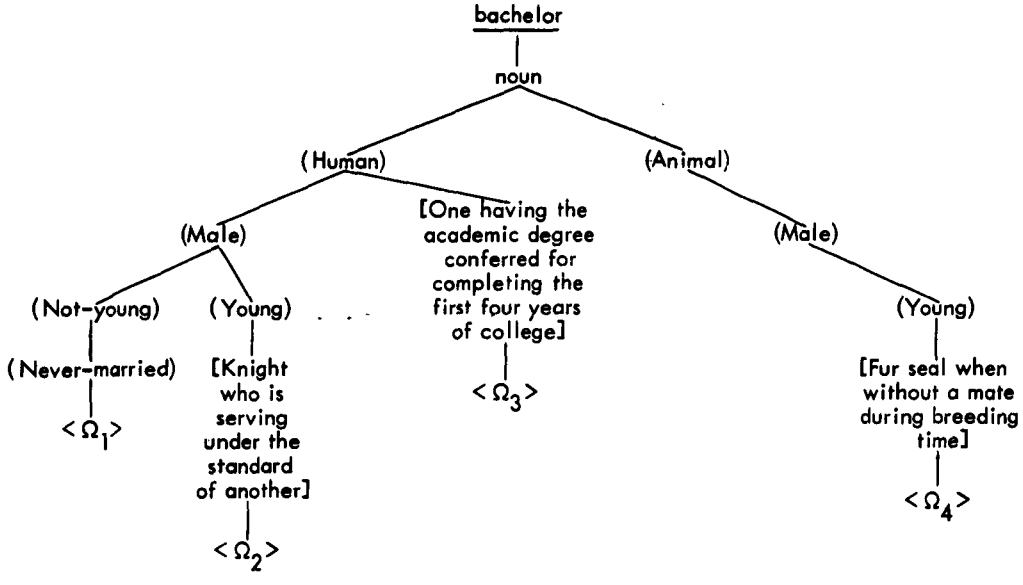


Fig. XIV-2.

of a lexical item. A distinguisher serves to distinguish a lexical item from those that are closest to it in meaning. Thus, a semantic marker found in the path of a certain lexical item will also be found in the paths of many other lexical items throughout the dictionary, whereas a distinguisher found in the path of a certain lexical item will, with very few exceptions, not be found anywhere else in the dictionary. This difference can be more fully appreciated if one compares the consequences of eliminating a semantic marker from a dictionary with the consequences of eliminating a distinguisher; in the former case indefinitely many semantic relations between the expressions of the language which were marked by the eliminated semantic marker would no longer be marked, whereas in the latter case only the few distinctions in sense which were marked by the eliminated distinguisher would no longer be marked.[12]

A lexical item is ambiguous if and only if its entry contains at least two distinct paths. It is clear, moreover, that ambiguity at the lexical level is the source of semantic ambiguity at the sentence level. Thus, a necessary, though not sufficient, condition for a syntactically unambiguous sentence to be semantically ambiguous is that it contain an ambiguous lexical item. For example, the source of the semantic ambiguity of the sentence He likes to wear a light suit in the summer is the lexical ambiguity of the word light. Since an adequate dictionary entry for a lexical item must mark every one of its ambiguities, the dictionary entry for light is required to represent this lexical item as branching into one path containing the semantic marker (Color) but not (Weight)

and another containing the semantic marker (Weight) but not (Color).

However, an ambiguous lexical item in a syntactically unambiguous sentence is not a sufficient condition for that sentence to be semantically ambiguous. For example, the sentence The stuff is light enough to carry, though it contains the ambiguous lexical item light, is not understood according to the sense in which light enough to carry means light enough in color to be carried. Thus, that the occurrence of an ambiguous lexical item in a sentence does not ipso facto make that sentence semantically ambiguous implies that the grammatical relations in the sentence and/or the meanings of the other constituents prevent this item from bearing more than one of its readings in its role as a constituent of the sentence. This shows that selection of some senses and exclusion of others occur as a result of the other constituents of the sentence. Such selection is of fundamental importance because, together with lexical ambiguity, it determines whether or not a sentence is anomalous, whether a sentence is semantically unambiguous or semantically ambiguous, and all other semantic properties of sentences which we want a semantic theory to mark.

Therefore, besides containing syntactic markers that determine the part-of-speech classification for a lexical item, semantic markers that represent the semantic properties that the lexical item shares with many other lexical items, and (optionally) a distinguisher that fixes its idiosyncratic features, a path for a lexical item can contain selection restrictions that determine the combinations into which the lexical item can enter and the sense(s) that it bears in those combinations. The formal representation of selection restrictions can be regarded as an explication of such information as The Shorter Oxford English Dictionary's qualification that the word honest when applied to persons means "of good moral character, virtuous, upright" and applied to women is ambiguous between this sense and the sense of chaste. In SST we wrote:

> For our reconstruction, we shall use left and right angles enclosing a Boolian function of syntactic or semantic markers. Such configurations of symbols will be affixed to the terminal element of a path (either the distinguisher or the last semantic marker if there is no distinguisher) and will be construed, relative to the projection rules, as providing a necessary and sufficient condition for a semantically acceptable combination. Thus, for example, the angle-material affixed to the path of a modifier determines the applicability of that path of the modifier to a sense of a nominal head. In particular, a path in the dictionary entry for honest will be: honest → adjective → (Evaluative) → (Moral) – [Innocent of illicit sexual intercourse] <(Human) & (Female)>. This is to be construed as saying that an adjectival occurrence of honest receives the interpretation (Evaluative) → (Moral) → [Innocent of illicit sexual intercourse] just in case the head it modifies has a path containing both the semantic marker (Human) and the semantic marker (Female). How in actual practice a semantic theory utilizes angle-material to determine selection-exclusion relations to obtain the correct semantic interpretations for sentences can only be made clear by a statement of the projection rules.[13]

The next notion for us to explain is the notion 'projection rule'. Let us suppose that an English grammar provides a semantic theory with the input sentence The boys like candy, together with the constituent structure characterization as given in Fig. XIV-1. The first step that the theory performs in the process of assigning a semantic interpretation to this sentence will be to associate with each of the lexical items in the sentence, i.e., the, boy, s, like, and candy, all and only the paths from their dictionary entries that are compatible with the syntactic categorization that the lexical items are given in the constituent structure characterization. This is the first point at which a significant use of grammatical information is made, but by no means the last. The semantic theory works as follows: a path in the dictionary entry for the lexical item m_j is assigned to the finite, non-null set of paths P_j^i which is associated with the occurrence of m_j in the constituent structure characterization d_i just in case that path contains syntactic markers that attribute to m_j the same part-of-speech classification that it has on the constituent structure characterization d_i. Thus, the lexical item m_1 is associated with the set of paths P_1^i; m_2 is associated with P_2^i, ..., and m_n is associated with P_n^i.[14] Referring to Fig. XIV-1, we picture the result of this step as converting the diagram into one in which the is associated with the set of paths P_1^i, boy is associated with P_2^i, s is associated with P_3^i, like is associated with P_4^i, and candy is associated with P_5^i (though no other change is made). Thus, for example, P_5^i contains paths representing each of the senses that candy has as a noun but none of the paths representing its senses as a verb, e.g., We will candy fruits tomorrow, The fruits candy easily, etc. This rule that associates senses with the occurrences of lexical items in constituent structure characterizations is thus our first projection rule. But since it is in many ways atypical, we shall continue our discussion of projection rules as if this were not such a rule.

There are two types of projection rules: type 1 projection rules and type 2 projection rules. The job of type 1 projection rules is to effect a series of amalgamations of paths, proceeding from the bottom to the top of a constituent structure characterization, by embedding paths into each other to form a new path, the amalgam. The amalgam is then assigned to the set of paths associated with the node that immediately dominates the sets of paths from which the amalgamated paths were drawn. The amalgam thus provides one of the ways to read the sequence of lexical items that the node dominates. In this manner, a set of readings is provided for every sequence of lexical items dominated by a syntactic marker in the constituent structure characterization, until the highest syntactic marker 'S' is reached and associated with a set of readings for the whole sentence. The operation of amalgamation is that of joining with one another one path from each of the n different sets of paths dominated by a syntactic marker SM to form a composite path to be a member of the set of paths associated with the node that the syntactic marker SM labels. The joining of a pair of paths occurs just in case one of the paths satisfies the selection restrictions that the other contains. If the syntactic

marker SM dominates just the sets of paths P_1^i, P_2^i, ..., P_n^i and P_1^i contains k_1 paths, P_2^i contains k_2 paths, ..., P_n^i contains k_m paths, then the set of paths which is associated with the dominating marker SM contains, at most $(k_1 \cdot k_2 \cdot \ldots \cdot k_m)$ members and possibly zero members if selection restrictions prevent every possible amalgamation from forming. Each path that is in the set assigned to SM is called 'a reading for the lexical string that SM dominates in the constituent structure characterization d_i'. The number of readings that is thus allotted to a string of lexical items determines its degree of semantic ambiguity: A string with no readings is anomalous, a string with exactly one reading is unambiguous, and a string with two or more readings is semantically ambiguous two or more ways.

An example of a projection rule of type 1 is:

(R1) Given two paths associated with nodes branching from the same node labelled SM, one of the form,

$\underline{\text{Lexical String}}_1 \rightarrow$ syntactic markers of head $\rightarrow (a_1) \rightarrow (a_2) \rightarrow \ldots \rightarrow (a_n) \rightarrow [1] <\Omega_1>$

and the other of the form,

$\underline{\text{Lexical String}}_2 \rightarrow$ syntactic markers of the modifier of the head $\rightarrow (b_1) \rightarrow (b_2) \rightarrow \ldots \rightarrow (b_m) \rightarrow [2] <\Omega_2>$

such that the string of syntactic or semantic markers of the head has a substring σ which satisfies $<\Omega_2>$, then there is an amalgam of the form,

$\underline{\text{Lexical String}}_2 + \underline{\text{Lexical String}}_1 \rightarrow$ dominating node marker SM $\rightarrow (a_1) \rightarrow (a_2) \rightarrow \ldots \rightarrow (a_n) \rightarrow (b_1) \rightarrow (b_2) \rightarrow \ldots \rightarrow (b_m) \rightarrow [[2][1]] <\Omega_1>$,

where any b_i is null just in case there is an a_j such that $b_i = a_j$, and $[[2][1]]$ is simply $[1]$ just in case $[2] = [1]$. This amalgam is assigned to the set of paths associated with the node labelled SM that dominates $\underline{\text{Lexical String}}_2 + \underline{\text{Lexical String}}_1$

(R1) explicates the process of attribution, i.e., the process of creating a new semantic unit compounded from a modifier and head whose semantic properties are those of the head, except that the meaning of the compound is more determinate than the head's by virtue of the semantic information contributed by the modifier. The erasure clause at the end of the statement of (R1) is included to avoid pointlessly duplicating the semantic markers and distinguishers in the path for a compound expression.[15] The modifier-head relation will be explicated by the grammar of the language so that the constituent structure characterization of a sentence will mark all cases of this relation that are found in it. In English, as well as many other languages, instances of modifier-head relations are: adjective-noun modification, adverb-verb modification, adverb-adjective modification, etc. Here, then, is another point at which grammatical information is utilized in the process of assigning a semantic interpretation to a sentence.

An example of an amalgamation produced by (R1) is the joining of the path colorful

→ adjective → (color) → [abounding in contrast or variety of bright colors] <(Physical object) v (Social activity)> and the path ball → noun → (Physical object) → [Of globular shape] to produce the new compound path colorful + ball → noun → (Physical object) → (color) → [[abounding in contrast or variety of bright colors][of globular shape]]. An example of an amalgamation that is prevented by a selection restriction is that of the path colorful → adjective → (Evaluative) → [of distinctive character, vividness, or picturesqueness] <(Aesthetic Object) v (Social Activity)> with the path for ball just given above. This possible amalgamation is precluded because the selection restriction in the path of the modifier requires that this path be joined only with paths of heads that contain either the semantic marker (Aesthetic object) or the semantic marker (Social activity), and this path of ball contains neither one of these semantic markers.

Type 2 rules work differently and are best explained after we explain the concept of a semantic interpretation of a sentence.

A semantic theory receives more than one constituent structure characterization for a sentence if that sentence is syntactically ambiguous. Figures XIV-3 and XIV-4 show the two constituent structure characterizations for the syntactically ambiguous sentence I like little boys and girls.

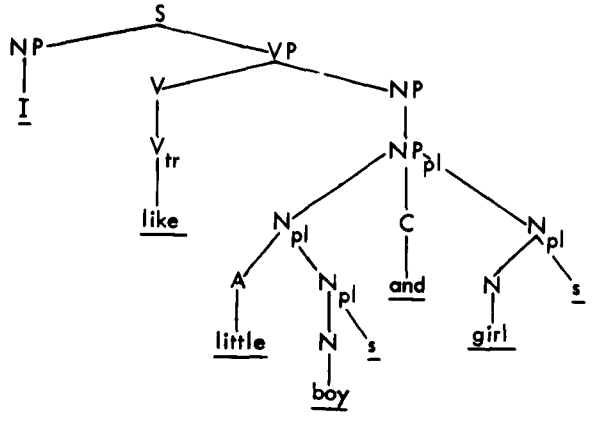


Fig. XIV-3.

Let $d_1, d_2, \ldots, d_n$ (for $n \geq 1$) be the constituent structure characterizations that the grammar provides for the sentence S. We shall define the "semantic interpretation of S" to be (1) the conjunction $\psi d_1 \varepsilon \psi d_2 \varepsilon \ldots \varepsilon \psi d_n$ of the semantic interpretations of the n constituent structure characterizations of S, and (2) the statements about S that

follow from the definition schema:

(D) S is fully X if and only if S is X on every d_i. The semantic interpretation ψd_i of the constituent structure characterization d_i of S is (1) the constituent structure characterization d_i each node of which is associated with its full set of readings, i.e., every reading that can belong to the set on the basis of the dictionary entries and the projection

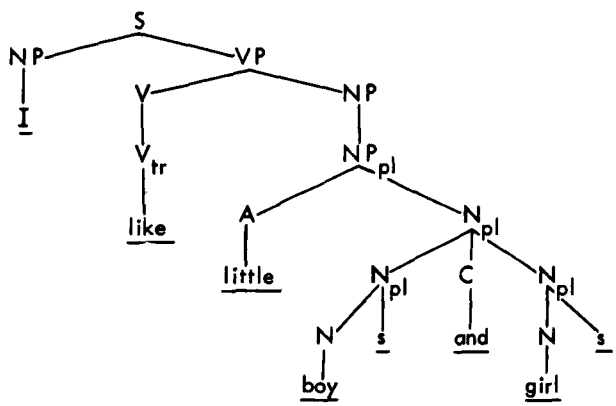


Fig. XIV-4.

rules does belong to it, and (2) the statements about S that follow from (1) together with the definitions:

(D1) S is semantically anomalous on d_i if and only if the set of paths associated with the node labelled 'S' in d_i contains no members.

(D2) S is semantically unambiguous on d_i if and only if the set of paths associated with the node labelled 'S' in d_i contains exactly one member.

(D3) S is n-ways semantically ambiguous on d_i if and only if the set of paths associated with the node labelled 'S' in d_i contains exactly n members ($n \geq 2$).

(D4) S_1 and S_2 are paraphrases on a reading with respect to their characterizations d_i and d_j if and only if the set of paths associated with the node labelled 'S' in d_i and the set of paths associated with the node labelled 'S' in d_j have a reading in common.

(D5) S_1 and S_2 are full paraphrases with respect to their characterizations d_i and d_j if and only if the set of paths associated with the node labelled 'S' in d_i and the set of paths associated with the node labelled 'S' in d_j have exactly the same membership.

Since these definitions are self-explanatory, we can now return to our account of the projection rules and explain the concept of a type 2 projection rule.

A grammar employs two types of syntactic rules to achieve its aim of assigning the

correct constituent structure characterization to each sentence of the language.[16] The first type are rules that develop single symbols by rewriting operations that are restricted to scanning the linear context of the symbol to be developed for information used to determine the rule's applicability. Such rules construct constituent structure characterizations such as those in Figs. XIV-1, XIV-3, and XIV-4 in somewhat the following manner: the first rule rewrites the initial symbol S (standing for "sentence") as NP+VP (which categorizes a noun phrase-verb phrase sequence as a sentence), then a rule can be used to rewrite NP as either NP_{sing} or NP_{pl}, then other rules can be used to rewrite VP as either V+NP or V_{intr} or be+Pred, still other rules to rewrite NP_{sing} as T+N, T as the, N as boy (or man, coat, mouse, etc.), and so forth.[17] But the scanning limitation on these rules makes the second type of rule necessary, for it has been shown that a grammar can assign constituent structure characterizations correctly only if some of its rules use information about the derivational history of sentences in order to determine their applicability.[18] Thus, in addition to such rewrite rules, grammars contain what are called "transformational rules," rules that operate on entire constituent structure characterizations, or any of their parts, and map labelled trees onto labelled trees, thus transforming simpler sentences into more complex ones and assigning the transformed sentence a constituent structure characterization. Transformational rules perform the task of explicating the syntactic relations between sentence types in the language. Thus, they show the syntactic relation between such pairs as The boys play games and Games are played by the boys, The men fly planes and What do the men fly?, etc., by showing how the latter member of each pair is constructed out of the former.

Type 2 projection rules are intended to explicate the manner in which transformational rules preserve or change meaning. It has often been observed by linguists who work with transformational rules that the sentence resulting from the application of a transformational rule to a set of source sentences is related in meaning to these source sentences in a definite, systematic way.[19] The employment of type 2 rules is intended to reveal the facts of language that underlie this observation.

We can characterize a type 2 projection rule as a rule that produces a semantic interpretation ψd_i for the constituent structure characterization d_i that has been constructed by the operation of the transformation T out of the set of constituent structure characterizations $d_1, d_2, \ldots, d_n$. A type 2 projection rule operates on the set of semantic interpretations $\psi d_1, \psi d_2, \ldots, \psi d_n$ and the transformation T to produce the semantic interpretation ψd_i. Type 2 projection rules should assign semantic interpretations in such a way as to reconstruct the manner in which the meaning of the sentence that was constructed by T is a function of the meanings of each of the sentences that were used by T in its construction.

J. J. Katz

Footnotes

1. A full discussion of the nature of a semantic meta-theory will be found in J. J. Katz and J. A. Fodor, The structure of a semantic theory (to be published in Language); Readings in the Philosophy of Language, edited by J. A. Fodor and J. J. Katz (Prentice-Hall, Inc., Englewood Cliffs, N.J., in press).

2. This is not meant to imply that the conception of a semantic theory which is outlined here conceives of such a theory as the product of a discovery procedure.

3. Here, I anticipate some such objection as the following: "How can you say a fluent speaker 'knows' something if he cannot say what it is that you claim he knows?" I do not think anything hangs on my having the word "know," and so if anyone insists upon this objection, I shall give up the word rather than become embroiled in a lexical quibble. I intend to convey the idea that the fluent speaker has acquired the necessary means for performing a task whose character compels us to admit that its performance results from the application of rules. Among the reasons that compel us to make this admission is that cited by Miller, Pribram, and Galanter, viz. that the task of understanding any twenty-word sentence is one that a fluent speaker can perform, yet the number of twenty-word sentences is 10^{30}, while the number of seconds in a century is only 3.15×10^{9}. Cf. G. A. Miller, K. Pribram, and E. Galanter, Plans and the Structure of Behavior (Henry Holt and Company, New York, 1960), pp. 146-147.

4. I say "in principle" because, in practice, limitations of perception, memory, mortality, etc., prevent the speaker from applying his knowledge of the rules of the language to provide himself with the meaning of certain sentences. This situation is exactly analogous to the case of a person's knowledge of the rules of arithmetic computation. Knowing how to perform any computation and knowing the rules of arithmetic computation are not sufficient to enable someone to actually perform any (specific) computation; for, again, limitations of perception, memory, mortality, etc., stand in the way.

5. Such information will be needed to provide the difference upon which rests the distinction in meaning between sentences composed of exactly the same morphemes, e.g., Gourmets do approve of people eating, Gourmets do approve of eating people, Do gourmets approve of people eating?, and so on.

6. N. Chomsky, Syntactic Structures (Mouton and Company, 'S-Gravenhage, Second Printing, 1962), Chapter 4. In general, we shall follow Chomsky's conception of syntax.

7. I shall use the notational abbreviations: "NP" for a noun phrase, "VP" for a verb phrase, "N" for a noun, "V" for a verb, "T" for an article, "A" for an adjective, "C" for a co-ordinating conjunction, and the subscript symbols "sing," "pl," and "tr" for the syntactic properties of nominal singularness, nominal pluralness, and verbal transitivity, respectively.

8. In particular, an optimal grammar will include a specification of the transformational history for each sentence. Cf. N. Chomsky, op. cit.

9. For the first of these two examples I am indebted to Professor Uriel Weinreich, Columbia University, and for the second to Professor George A. Miller, Center for Cognitive Studies, Harvard University.

10. In a small minority of cases, dictionary entries consist of instructions, e.g., the rules for not that are given in J. J. Katz, Analyticity and contradiction in natural language, Readings in the Philosophy of Language, op. cit. For a further discussion of the type of entry found in the vast majority of cases see J. J. Katz and J. A. Fodor, op. cit., Section 6.

11. Two comments on Fig. XIV-2. First, the word bachelor, though a noun, can select and exclude other nouns in various types of constructions, e.g., in noun-noun cases such as He is my bachelor friend, or in noun-in-apposition cases such as Mr. Smith, the neighborhood bachelor, is here. Thus, we must represent bachelor as having a selection restriction for each sense; thus the terminal element for each path in Fig. XIV-2 is a selection restriction enclosed in angles. Second, the particular selection restrictions are omitted because their inclusion would only complicate matters unnecessarily at this point.

12. For a further examination of the distinction between the notions 'semantic marker' and 'distinguisher', see J. J. Katz and J. A. Fodor, op. cit., Section 6.

13. Ibid., loc. cit.

14. For a full discussion of this step see the treatment of rule (I) in Katz and Fodor, ibid., Section 7.

15. Thus, for example, it makes no sense to include the semantic markers (Human) and (Female) twice in the path associated with the compound spinster aunt because both of the constituent paths contain occurrences of both. The second occurrence of (Human) or (Female) would provide no semantic information whatever.

16. Although this is not the only aim of a grammar. Cf. N. Chomsky, On the notion 'rule of grammar', Proc. Symposia in Applied Mathematics, Vol. 12, The Structure of Language and Its Mathematical Aspects, edited by R. Jakobson (American Mathematical Society, New York, 1961), pp. 6-24.

17. Cf. N. Chomsky, A transformational approach to syntax, Third Texas Conference on Problems of Linguistic Analysis of English, edited by A. A. Hill (University of Texas, 1962), pp. 123-158.

18. N. Chomsky, Syntactic Structure, op. cit.; On the notion 'rule of grammar', op. cit.; The Logical Structure of Linguistic Theory, 1955 (microfilm available in Hayden Library, M.I.T.). Cf. P. Postal, Some further limitations of phrase-structure grammars, Readings in the Philosophy of Language, op. cit.

19. This point has been discussed outside the context of the conception of a semantic theory adopted in the present report in two recent papers: J. A. Fodor, Projection and paraphrase in semantics, Analysis 21, 73-77 (1961); J. J. Katz, A reply to 'Projection and Paraphrase in Semantics', Analysis 22, 36-41 (1961).

B. REMARKS ON THE MORPHOPHONEMIC COMPONENT OF RUSSIAN

In this report we present an entirely new synchronic analysis of the morphophonemic component of the contemporary standard literary dialect of Russian. Although our analysis is original, we nevertheless draw the reader's attention to the general similarity of the rules that we propose here and the rules already proposed by previous investigators.[1] The originality of our analysis, therefore, consists primarily in the underlying (phonemic) structure which we postulate for Russian.

We consider here only verb forms, and in the rules presented below we shall not dwell on particularities of pronunciation which have been adequately described elsewhere.[2]

We shall be concerned in particular with the following forms:

	Infinitive	1 Sg	3 Sg	3 Pl	Imperative	Masc Past
1.	n,ɪst,ʃ	n,ɪsú	n,ɪs,ɔ́t	n,ɪsút	n,ɪs,ʃt,ɪ	n,ɔ́s
2.	t,ɛ́č,	t,ɪkú	t,ɪč,ɔ́t	t,ɪkút	t,ɪk,ʃt,ɪ	t,ɔ́k
3.	l,ub,ʃt,	l,ubl,ú	l,úb,ɪt	l,úb,ɪt	l,ub,ʃt,ɪ	l,ub,ʃl
4.	plák⊖t,	pláč,u	pláč,ɪt	pláč,ut	pláč,t,ɪ	plák⊖l
5.	l,ɪžát,	l,ɪžú	l,ɪžát	l,ɪžát	l,ɪžát,ɪ	l,ɪžál
6.	znát,	znáju	znájɪt	znájut	znájt,ɪ	znál
7.	žát,	žávú	žáv,ɔ́t	žávút	žáv,ʃt,ɪ	žál
8.	abrəzavát,	abrazúju	abrazújɪt	abrazújut	abrazújt,ɪ	abrəzavál
9.	gl,ɪd,ɛ́t,	gl,ɪžú	gl,ɪd,ʃt	gl,ɪd,ɛ́t	gl,ɪd,ʃt,ɪ	gl,ɪd,ʃl

We assume the syntactic component of the grammar to produce the forms presented below. Parentheses indicate immediate constituent structure. We consider our task to be that of deriving the phonetic representations above from the phonemic representations presented below:

	Infinitive	1 Sg	3 Sg	3 Pl	Imperative	Masc Past
1.	(nes+tI)	((nes+e)+ou)	((nes+e)+t)	((nes+e)+out)	(((nes+e)+I+#)+te)[3]	(nes+l)
2.	(tek+tI)	((tek+e)+ou)	((tek+e)+t)	((tek+e)+out)	(((tek+e)+I+#)+te)	(tek+l
3.	(leub+I+tI)	((leub+I+I)+ou)	((leub+I+I)+t)	((leub+I+I)+ēt)	(((leub+I+I)+I+#)+te)	(leub+I+l)
4.	(plōk+ō+tI)	((plōk+ō+e)+ou)	((plōk+ō+e)+t)	((plōk+ō+e)+out)	(((plōk+ō+e)+I+#)+te)	(plōk+ō+l)
5.	(leg+ē+tI)	((leg+ē+I)+ou)	((leg+ē+I)+t)	((leg+ē+I)+ēt)	(((leg+ē+I)+I+#)+te)	(leg+ē+l)
6.	(znōi+tI)	((znōi+e)+ou)	((znōi+e)+t)	((znōi+e)+out)	(((znōi+e)+I+#)+te)	(znōi+l)
7.	(gIu+tI)	((gIu+e)+ou)	((gIu+e)+t)	((gIu+e)+out)	(((gIu+e)+I+#)+te)	(gIu+l)
8.	(obrōz+ou+ō+tI)	((obrōz+ou+ō+e)+ou)	((obrōz+ou+ō+e)+t)	((obrōz+ou+ō+e)+out)	(((obrōz+ou+ō+e)+I+#)+te)	(obrōz+ou+ō+l)
9.	(glēd+ei+tI)	((glēd+ei+I)+ou)	((glēd+ei+I)+t)	((glēd+ei+I)+ēt)	(((glēd+ei+I)+I+#)+te)	(glēd+ei+l)

One of the Morpheme-Structure Rules that we postulate for Russian states that no terminal vocabulary item may end in a vowel. This rule will convert terminal vocabulary items like /znōi/ and /gīu/ to /znōj/ and /gīw/. For the purposes of the present report we may state this MS-Rule as follows:

MS-1 [-cons] → [-voc] in env: _____ +

Note that because of this rule we may use archiphonemes specified solely for the features consonantal and gravity in terminal vocabulary items like /znōI/ and /gīU/, etc.

In the phonemic forms presented above we have used the following vowel system:

segment:	ū	ī	ō	ē	u	i	o	e
tense:	+	+	+	+	-	-	-	-
diffuse:	+	+	-	-	+	+	-	-
grave:	+	-	+	-	+	-	+	-

We use consonant letters as abbreviations for the appropriate distinctive feature matrices. We draw attention, however, to the fact that in the forms presented above palatalization (sharping) of consonants is completely predictable (this fact may be formally presented in a Morpheme-Structure Rule to the effect that all consonants – with some limitations which we have discussed elsewhere – are specified non-sharp).

We require application of the following rules to the phonemic representation in order to derive the correct phonetic representations. C-Rules apply cyclically to segments within innermost parentheses; P-Rules apply to forms derived from the C-Rules.

C-1 Insert j in env: _____ + $\begin{bmatrix}+\text{voc}\\-\text{cns}\\+\text{tns}\end{bmatrix}$ + $\begin{bmatrix}+\text{voc}\\-\text{cns}\\-\text{tns}\end{bmatrix}^4$

C-2 u → w in env: _____ + $\begin{bmatrix}+\text{voc}\\-\text{cns}\end{bmatrix}$

C-3 V_1^2 → Ø in env: _____ + $\begin{bmatrix}+\text{voc}\\-\text{cns}\end{bmatrix}$

C-4 Erase parentheses and return to C-1. If there are no more parentheses, then go to P-1.

P-1 $\begin{bmatrix}+\text{obs}\\+\text{cmp}\end{bmatrix}$ → $\begin{bmatrix}+\text{str}\\-\text{grv}\end{bmatrix}$ in env: _____ $\begin{bmatrix}+\text{voc}\\-\text{cns}\\-\text{grv}\end{bmatrix}$ X

where X is not + # .

P-2 [+cons] → [+sharp] in env: _____ $\begin{bmatrix}-\text{cns}\\-\text{grv}\end{bmatrix}$

P-3 I → Ø in env: X + (t,) _____ + #

where X contains a stressed vowel and does not end in two consonants.

P-4 Transitive softening (i.e., s,j → š, , d,j → ǰ, , b,j → bl, , k,j → č, , etc., and j → Ø in env: C _____).

P-5 e → o in env: _____ [+cons, -shrp]

P-6 X → [-cmp, -flt]

P-7 [-tns, +grv] → [+flt]

P-8 [-dif] → [+cmp]

} in env: [+voc, -cns]

P-9 ɛ ɪ → ɪ ɛ

P-10 V → Ø in env: _____ V

P-11 w → v

P-12 { š, → š ; ǰ, → ž ; č, → č }

P-13 æ → a

P-14 Vowel reduction; raising of [ɛ́] to [é] before soft consonants; backing of [i] to [ɨ] after hard consonants.

P-15 [-voc, -cns] → Ø in env: _____ + [+cons]

P-16 {d, t} → Ø in env: _____ + l

P-17 l → Ø in env: C + _____

P-18 Erase all boundary and juncture markers.

P-Rules 6-8 apply to all vowels and specify features absent in the matrix presented above. Note that after application of P-Rules 6-8 the vowel segments presented above will have the following distinctive feature matrices:

segment:	ū	ī	ō	ē	u	i	o	e
tense:	+	+	+	+	-	-	-	-
diffuse:	+	+	-	-	+	+	-	-
grave:	+	-	+	-	+	-	+	-
compact:	-	-	+	+	-	-	+	+
flat:	-	-	-	-	+	-	+	-

Inspection of this matrix will reveal that application of P-Rules 6-8 have the following effect (symbols to the left of the arrow those of the matrices above, to the right of the

arrow those of the International Phonetic Alphabet):

ū	→	ɨ		u	→	ʊ
ī	→	i		i	→	ɪ
ō	→	a		o	→	ɔ
ē	→	æ		e	→	ɛ

We apply these rules to some of the phonemic forms presented above:

1a: (nes+tī) →C-4→ nes+tī →P-2→ n,es+t,ī →P-14→ n,ɪs+t,í →P-18→ n,ɪst,í

1b: ((nes+é)+t) →C-4→ (nes+é+t) →C-4→ nes+é+t →P-2→ n,es,+é+t →P-5→ n,es,+ó+t →P-6, P-7, P-8→ n,ɛs,+ɔ́+t →P-14→ n,ɪs,+ɔ́+t →P-18→ n,ɪs,ɔ́t

1c: (nés+1) →C-4→ nés+1 →P-2→ n,és+1 →P-5→ n,ós+1 →P-6, P-7, P-8→ n,ɔ́s+1 →P-17→ n,ɔ́s

2a: (ték+tī) →C-4→ ték+tī →P-2→ t,ék+t,ī →P-3→ t,ék+t, →P-4→ t,éč, →P-6, P-8, P-14→ t,éč,

2b: ((tek+é)+t) →C-4→ (tek+é+t) →C-4→ tek+é+t →P-1→ teč+é+t →P-2→ t,eč,+é+t →P-5→ t,eč,+ó+t →P-6, P-7, P-8, P-14→ t,ɪč,+ɔ́+t →P-18→ t,ɪč,ɔ́t

2c: (((tek+e)+ī+#)+te) →C-4→ ((tek+e+ī+#)+te) →C-3→ ((tek+ī+#)+te) →C-4→ (tek+ī+#+te) →C-4→ tek+ī+#+te →P-2→ t,ek,+ī+t,e →P-6, P-8, P-14, P-18→ t,ɪk,ít,ɪ

3a: ((leub+ī+ĭ)+ou) →C-3→ ((leub+ĭ)+ou) →C-4→ (leub+ĭ+ou) →C-1→ (leubj+ĭ+ou) →C-3→ (leubj+ou) →C-4→ leubj+ou →P-2→ l,eub,j+ou →P-4→ l,eubl,+ou →P-10→ l,ubl,+ú →P-18→ l,ubl,ú

4a: ((plōk+ō+e)+t) →C-1→ (plōkj+ō+e)+t) →C-3→ ((plōkj+e)+t) →C-4→ (plōkj+e+t) →C-4→ plōkj+e+t →P-2→ plōk,j+e+t →P-4→ plōč,+e+t →P-5→ plōč,+o+t →P-6, P-7, P-8→ pláč,+ɔ+t →P-14→ pláč,+ɪ+t →P-18→ pláč,ɪt

4b: (((plōk+ō+e)+ī+#)+te) →C-1→ (((plōkj+ō+e)+ī+#)+te) →C-3→ (((plōkj+e)+ī+#)+te) →C-4→ ((plōkj+e+ī+#)+te) →C-3→ ((plōkj+ī+#)+te) →C-4→ (plōkj+ī+#+te) →C-4→ plōkj+ī+#+te →P-2→ plōk,j+ī+#+t,e →P-3→ plōk,j+#+t,e →P-4→ plōč,+#+t,e →P-6, P-8, P-14, P-18→ pláč,t,ɪ

5a: (leg+ē+1) →C-4→ leg+ē+1 →P-1→ leǰ+ē+1 →P-2→ l,eǰ,+ē+1 →P-6, P-8→ l,ɛǰ,+æ+1 →P-12→ l,ɛž+æ+1 →P-13→ l,ɛž+á+1 →P-14→ l,ɪž+á+1 →P-18→ l,ɪžál

6a: (znōj+tī) →C-4→ znōj+tī →P-2→ znōj+t,ī →P-3→ znōj+t, →P-6, P-8→ znáj+t, →P-15→ zná+t, →P-18→ znát,

6b: (((znōj+e)+ī+#)+te) →C-4→ ((znōj+e+ī+#)+te) →C-3→ ((znōj+ī+#)+te) →C-4→

(znōj+Ī+#+te) →C-4→ znōj+ī+#+te →P-2→ znōj+ī+#+t,e →P-3→ znōj+#+t,e →P-6, P-8→ znáj+#+t,ɛ →P-14→ znáj+#+t,I →P-18→ znájt,I

8a: ((obrōz+ou+ō+e)+ou) →C-1→ ((obrōz+ouj+ō+e)+ou) →C-3→ ((obrōz+ouj+e)+ou) →C-4→ (obrōz+ouj+e+ou) →C-3→ (obrőz+ouj+ou) →C-4→ obrōz+ouj+ou →P-6, P-7, P-8→ ɔbraz+ɔuj+ɔu →P-10→ ɔbraz+új+u →P-14→ abraz+új+u →P-18→ abrazúju

8b: (obrōz+ou+ō+l) →C-2→ (obrōz+ow+ō+l) →C-4→ obrőz+ow+ō+l →P-6, P-7, P-8→ ɔbraz+ɔw+á+l →P-11→ ɔbraz+ɔv+á+l →P-14→ abrəz+av+á+l →P-18→ abrəzavál

9a: (glēd+ei+l) →C-4→ glēd+ei+l →P-2→ gl,ēd,+ei+l →P-6, P-8→ gl,æd,+ɛI+l →P-9→ gl,æd,+Iɛ+l →P-10→ gl,æd,+é+l →P-14→ gl,Id,+é+l →P-18→ gl,Id,él

T. M. Lightner

References

1. We refer in particular to Roman Jakobson, Russian conjugation, Word 4, 155-167 (1948), M. Halle, Note on cyclically ordered rules in the Russian conjugation, Quarterly Progress Report No. 63, Research Laboratory of Electronics, M.I.T., October 15, 1961, pp. 149-155, T. M. Lightner, On pon,át, and obrazovát, type verbs in Russian, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 177-180, G. H. Matthews and T. M. Lightner, On the Present tense theme e/o in Russian, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M.I.T., January 15, 1963, pp. 190-193.

2. See, e.g., rules P-4, P-9, P-10, P-12, P-14. Details on these rules may be found in the works cited above. Information on stress prediction can be found in Jakobson, op. cit., and Halle, op. cit.; in this report we simply mark the stress where it falls and do not attempt to add anything to the work of previous investigators. For a more detailed discussion on the monophthongization of Slavic diphthongs (condensed here in rules P-9 and P-10), see Roman Jakobson, Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slaves, III et passim, Travaux du Cercle Linguistique de Prague (Jednota Československých Matematiků a Fysiků, Prague, 1929), and more recently, M. Halle, The Proto-Slavic diphthongs, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 296-297.

3. For the necessity of assuming a # juncture in the imperative, see Roman Jakobson, Russian conjugation, op. cit., and M. Halle, The Sound Pattern of Russian (Mouton and Company, The Hague, 1959), p. 67. Note that this rule will prevent application of P-15 to forms like /znáj+#+t,I/ (ultimately [znájt,I]).

4. We must place a restriction on the application of rule C-1. The restriction formulated by Halle in his "Note on cyclically ordered rules in the Russian conjugation" will not satisfy us because we do not consider ždat', lgat', brat', etc. to be formed from nonsyllabic roots, but rather from the roots /gid/, /lug/, /bir/, etc. In order to prevent application of C-1 to these verbs, we intend to follow some variant of Halle's most recent suggestion that this information be listed in the Complex Symbol for monosyllabic roots containing the archiphoneme {i,u}. Note that we require the presence of this vowel not only in order to account for the derived imperfectives vyzidat', oblygat', vybirat', etc., but also to account for the derived nominals lož', vybor, etc. (we defer treatment of imperfective derivation and nominalization for a later report). We shall postulate a special marker in the Complex Symbol to signal retention of these lax, diffuse vowels in the present tense of brat' (beru), zvat' (zovu), drat' (deru), stlat' (stelju).

We require further that the verb sosat' (sosu, sosët, etc.) be derived from the root /sus/, not only to account for the lack of application of C-1 but also to account for the dialectal forms ssu, ssët, etc. See A. G. Preobraženskij, Etimologičeskij slovar' russkago jazyka (Columbia University Press, New York, 1951), II, 360, and Max Vasmer, Russisches etymologisches Wörterbuch (Carl Winter, Universitätsverlag, Heidelberg, 1953), II, 701.

C. TYPE 1 GRAMMARS AND LINEAR-BOUNDED AUTOMATA

Recently, Landweber[1] showed that the language accepted by a deterministic linear-bounded automaton, in the sense of Myhill,[2] can be generated by a type 1 grammar, in the sense of Chomsky.[3] Landweber's proof remains valid for a nondeterministic linear-bounded automaton. As the converse of Landweber's theorem, we have

THEOREM: The language generated by a type 1 grammar is accepted by a nondeterministic linear-bounded automaton.

Thus, we now have equivalent heirarchies: Turing machines, linear-bounded automata, pushdown storage, and finite automata on the one hand, and semi-Thue systems, type 1 grammars, type 2 grammars, and type 3 grammars, on the other. (For the first pair, see, for example, Davis[4]; for the last two pairs, Chomsky.[3, 5])

The proof of the theorem will consist of three lemmas. Before stating the lemmas, we shall define a few notions. According to Chomsky,[3] a grammar is type 1 if each rule is of the following shape:

$$\phi A \psi \rightarrow \phi \omega \psi, \; l(A) = 1, \; l(\omega) \neq 0.$$

Here, $l(\phi)$ means the length of ϕ. We generalize his notion a bit, and understand a type 1 grammar to be a semi-Thue system in which each of its rules $\phi \rightarrow \phi'$ satisfies $l(\phi) \leq l(\phi')$. If, furthermore, (a) $l(\phi) = l(\phi')$ and ϕ' does not contain the S-symbol S, or (b) $\phi = S$, then the grammar is called length-preserving. On the other hand, we define the order of a grammar to be the maximum of the lengths of the strings appearing in the rules. If an order 2 grammar is length-preserving and if $S \rightarrow EF$ implies that $E = S$, it is said to be linear-bounded. Then we have the following lemmas.

LEMMA 1: For any type 1 grammar G, there exists an order 2 type 1 grammar G' equivalent to G: $L(G) = L(G')$.

Here, L(G) means the language generated by G. Notice that, since an order 2 type 1 (in the above-mentioned sense) grammar is easily seen to be equivalent to an order 2 grammar that is type 1 in Chomsky's sense, our notion of type 1 grammar turns out to be equivalent to the notion of type 1 grammar in Chomsky's sense, as far as the equivalence of grammars is concerned.

LEMMA 2: For any order 2 type 1 grammar G', there exists a linear-bounded grammar G'' equivalent to G': L(G') = L(G'').

LEMMA 3: For any linear-bounded grammar G'', there exists a nondeterministic linear-bounded automaton that accepts L(G'').

Our theorem follows directly from these lemmas.

S.-Y. Kuroda

References

1. P. S. Landweber, Three theorems on phrase structure grammar of type 1, (unpublished).

2. J. Myhill, Linear Bounded Automata, WADD Technical Note No. 60-165, Wright-Patterson Air Force Base, Ohio, 1960.

3. N. Chomsky, On certain formal properties of grammars, Information and Control 2, 137-167 (1959).

4. M. D. Davis, Computability and Unsolvability (McGraw-Hill Book Company, New York, 1958).

5. N. Chomsky, Context-free grammars and pushdown storage, Quarterly Progress Report No. 65, Research Laboratory of Electronics, M.I.T., April 15, 1962, pp. 187-194.

D. FORMAL JUSTIFICATION OF VARIABLES IN PHONEMIC CROSS-CLASSIFYING SYSTEMS

Variables in phonemic rules have proved extremely useful.[1-3] There is also a clear formal necessity that variables be included in any system with cross-classificatory features (whether or not the features are binary). If variables were not cost-free with respect to a simplicity criterion, the segments [+compact], [+grave] would appear to be related to each other in a more fundamental sense than the segments [-compact], [+grave]. But it is only an arbitrary decision to measure frontness in terms of gravity. '+grave' is exactly equivalent to '-acute': the value of the polarity of a given feature quality is not a substantive part of the theory. Therefore the metatheory requires that the features in a cross-classifying system are all marked:

$$[O\,\alpha F]$$

$$F \rightarrow \begin{matrix} \text{grave} \\ \text{compact} \\ \vdots \end{matrix}$$

$$\alpha \rightarrow \begin{Bmatrix} + \\ - \end{Bmatrix}$$

$$O \rightarrow \begin{Bmatrix} e \\ \sim \end{Bmatrix} \quad \text{(by convention, the identity operation e is left blank)}$$

with the operations $\sim + = -$ $\quad e+ = +$
$\sim - = +$ $\quad e- = -$

In phonemic matrices and nonvariable rules the value of O is usually affirmative, and α is specified with '+' or '-', but this is purely a convention; it would be exactly equivalent to maintain generally the value of $\alpha = +$ and specify the phonemic matrices by the symbols '$\sim$' or 'e'.

The availability of different operators clearly shows that the segments [+compact], [-grave] are in as close a relation as [+compact], [+grave]. For instance, the assimilation rules

$$[\ \ ] \rightarrow [+\text{comp}] \text{ in the env } \_\_ [+\text{grave}]$$
$$\rightarrow [-\text{comp}] \text{ in the env } \_\_ [-\text{grave}]$$

are more simply combined

A) $[\ \ ] \rightarrow [\alpha\text{comp}] \text{ in the env } \_\_ [\alpha\text{grave}]$

and the assimilation rules

$$[\ \ ] \rightarrow [-\text{comp}] \text{ in the env } \_\_ [+\text{grave}]$$
$$\rightarrow [+\text{comp}] \text{ in the env } \_\_ [-\text{grave}]$$

are combined

B) $[\ \ ] \rightarrow [\sim\alpha\text{comp}] \text{ in the env } \_\_ [\alpha\text{grave}]$.

If the front quality of segments were marked with acuteness instead of gravity, rules A) and B) would be equally simple: only the polarity is changed

A') $[\ \ ] \rightarrow [\sim\alpha\text{comp}] \text{ in the env } \_\_ [\alpha\text{acute}]$
B') $[\ \ ] \rightarrow [\alpha\text{comp}] \text{ in the env } \_\_ [\alpha\text{acute}]$.

The arbitrariness of the polarity values is represented by the existence of cost-free variables that make $[+X] = [\sim -X']$ where X and X' are opposite extreme values along the same feature continuum (e.g. {acute, grave; abstract, concrete}).

In phonology the use of variables over + and - and operators covers various phenomena:

<u>assimilation</u>
$[\ \ ] \rightarrow [\alpha X] \text{ in the env } \_\_ [\alpha X]$

<u>dissimilation</u>
$[\ \ ] \rightarrow [\sim\alpha X] \text{ in the env } \_\_ [\alpha X]$

<u>internal assimilation</u>
$[\alpha X] \rightarrow [\alpha Y]$

<u>internal dissimilation</u>
$[\alpha X] \rightarrow [\sim\alpha Y]$

external environment specification

$$[\quad] \rightarrow [+X] / \text{———} [\alpha X] [\sim \alpha Y]$$

exchange (internal dissimilation)

$$[\alpha X] \rightarrow [\sim \alpha X]$$

If the exchange rule applies in a transformational cycle, the net effect is one of reciprocation with respect to the feature X. If the affected segment is contained within an odd number of constituents to which the exchange rule applies, the net effect is dissimilative; if it applies an even number of times, there is no net effect. (Sections XIV-E, XIV-F, and XIV-G present the use of such a rule in the Indo-European e/o ablaut.)

T. G. Bever

References

1. M. Halle, A descriptive convention for treating assimilation and dissimilation, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M. I. T., July 15, 1962, pp. 295-296; M. Halle, The Proto-Slavic diphthongs, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 296-297.

2. T. G. Bever, Theoretical implications of Bloomfield's "Menomini Morphophonemics," Quarterly Progress Report No. 68, Research Laboratory of Electronics, M.I.T., January 15, 1963, pp. 197-203.

3. T. M. Lightner, Vowel harmony in classical (literary) Mongolian, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M.I.T., January 15, 1963, pp. 189-190.

E. THE RECIPROCATING CYCLE OF THE INDO-EUROPEAN E/O ABLAUT

The Indo-European (IE) e/o ablaut can be described by a reciprocating rule of the type discussed in Section XIV-D. The traditional presentation of the e/o ablaut is in terms of the cases, tenses or other derivations of the ablauting stems. An extremely telling observation is that words in compounds often have the opposite grade from the words alone.[1] This clearly indicates that the number of constituents in which the ablauting stem is contained is critical. In generative grammar, the combination of a morphophonemic cycle and an exchange rule is sensitive to the odd or even quality of the number of constituents. The IE rule is of the form

$$[\alpha\text{grave}] \rightarrow [\sim \alpha\text{grave}]$$

and it applies in a cycle. If the number of constituents containing the ablauting vowel is odd, the grade is changed; if it is even, the grade is unchanged. Sections XIV-F and XIV-G present the operation of this rule in Germanic and Greek. In the Germanic languages the scope of the ablaut is sharply restricted, but the similarity of the essential rule to that of Greek indicates that IE itself had a morphophonemic transformational

cycle that generated the e/o ablaut alternations.

T. G. Bever, D. T. Langendoen

References

1. J. Kurylowicz, L'apophonie en Indo-Européen (Polska Akademia Nauk, Warsaw, 1956), p. 71.

F. THE E/O ABLAUT IN OLD ENGLISH

The Germanic reflex of the Indo-European e/o ablaut appears in the nonreduplicating strong stems. The Germanic strong verbs, by Grimm's definition, show stem-vowel changes in derived forms. This analysis follows Keyser's suggestion that the historical distinction between the strong and weak stems is operative in Old English: strong stems are phonemically monosyllabic; weak stems, polysyllabic.[1]

The Germanic nonreduplicating verb ablaut occurs in 6 classes that are traditionally presented as 6 "ablaut series," each containing four forms. Since each of the classes is phonemically distinct, the entire set of series can be generated by 5 basic rules. Each rule is concerned with a single change in quality or quantity. This set occurs in a morphophonemic transformational cycle.

The traditionally presented "principal parts" of the old Germanic strong verb are: present infinitive, preterite singular, preterite plural, and past participle.

1. Past Participle

The strong verb past participle throughout old Germanic is generated by the rules

$$\begin{bmatrix}+\text{voc}\\-\text{cons}\end{bmatrix} \quad \begin{cases} \rightarrow \emptyset \text{ in the env } \_\_\_\_ \begin{bmatrix}+\text{voc}\\-\text{cons}\end{bmatrix} \\ \rightarrow [+\text{grave}] \text{ in the env } \_\_\_\_ [+\text{sonorant}] \end{cases}$$

Although these rules can be combined with the strong stem cycle, they will be omitted to simplify this presentation.

The relevant constituent structure is described by the rules

$$\text{Verb} \rightarrow \text{stem (+extension)} - \text{Mood}$$

$$\text{Mood} \rightarrow \text{PerNo} + \begin{cases}\text{indicative}\\ \text{subjunctive}\\ \text{infinitive}\\ \text{participle}\end{cases}$$

$$\text{PerNo} \rightarrow \begin{Bmatrix}1\\2\\3\end{Bmatrix} \begin{Bmatrix}\text{Sg}\\ \text{Pl}\end{Bmatrix}$$

$$\text{Extension} \rightarrow \left\{\begin{matrix}+\text{ substantive}\\ +\text{ past}\end{matrix}\right\}$$

$$\left.\begin{matrix}\left\{\begin{matrix}1\\3\end{matrix}\right\} & \text{sg} \rightarrow +\emptyset\\ \left\{\begin{matrix}1\\2\\3\end{matrix}\right\} & \text{pl} \rightarrow +\text{ on}\end{matrix}\right\} \quad \text{in the env past} + \text{———}$$

In Germanic, the ablauting of strong stems is restricted to the past tense or the derivation of nouns. This is represented by the optional stem extension: the cyclic rules will apply to an extended strong stem only. The actual morphophonemic value of the /+past/ stem extension is not critical, although it is probably $\emptyset$. This interpretation is not proposed as an ultimate solution: it merely represents the fact that the occurrence of the ablaut is limited in Germanic.

The form of the stems to which the stem cycle applies is

$$(((\#\ \text{stem} + \text{Ext}) + \left\{\begin{matrix}\text{Sg}\\ \text{Pl}\end{matrix}\right\}) + \#)$$

and in the cases presented here

$$(((\#\ \text{stem} + \text{Ext}) + \left\{\begin{matrix}\emptyset\\ \text{on}\end{matrix}\right\}) + \#),$$

where /+ext/ has the value /+past/. In OE the rule

C-1) $\quad [\alpha\text{grave}] \rightarrow [\sim\alpha\text{grave}]$ in the env $\left[\overline{\text{Long Syllable}}\right]$

is a reciprocating rule of the type discussed in Section XIV-E. Among the Germanic languages this rule is restricted to OE because only in OE does the distinction between /æ/ and /a/ "re-emerge" from reconstructed Proto-Germanic.

The environment "Long Syllable" is the same as that formulated by Keyser for the weak verb cycle[1]:

$$[+\text{cons}]\ [+\text{voc}]_{\alpha}\ [+\text{son}]_{\beta}\ [+\text{cons}],$$

where $\alpha + \beta \geq 2$.

The cycle reflects the historical verb class derivation,[2,3] since it is fairly clearly divided into two sections: the first 3 rules apply critically to the first 3 classes and the final 2 rules apply critically to the last 3 classes. The fact that all the environments are included between $\#C_0^3$ ——— C blocks the application of rule C-1) to the vowel in /rid/ or /far/. It also blocks the application of rule C-4) to each vowel of /raad/ or the vowel of /band/ because rule C-4) is actually

$$[+\text{comp}]_1^1 \rightarrow [+\text{long}]_1^1 \text{ in the env } \#C_0^3 \text{——} C + \text{ext} \left\{\begin{matrix}\emptyset\\ +\text{on}\end{matrix}\right.$$

The effect of the environment /+ext $\emptyset$/ is to limit the application of rules C-4-b) and C-5) to the first cycle, when nothing follows the /+ext/.

One incorrect form is generated by these rules – the preterite singular of /beeodon/.

It comes out /baaæd/ instead of the correct form /bææad/, but the diphthong structure rules will correct this. In order not to violate the distinctness convention the diphthong structure rules need to be included at the beginning of the cycle.

The Old English Weak Verb Cycle

		Pre-final restriction	Final restriction	
C-1)	$[\alpha\text{grave}] \rightarrow [\sim\alpha\text{grave}]$ in the env	$\longleftrightarrow$ [+son] $\longleftrightarrow$	no restriction	C-1
C-2)	$[\quad] \rightarrow$ [+diff] in the env	$\longleftrightarrow$ [+son] $\longleftrightarrow$	$+\begin{bmatrix}+\text{voc}\\-\text{cons}\end{bmatrix}_1$	C-2
C-3)	$[\quad]_1^n \rightarrow \emptyset$ in the env	$\underline{\quad}\begin{bmatrix}+\text{voc}\\-\text{cons}\end{bmatrix}$	$+\begin{bmatrix}+\text{voc}\\-\text{cons}\end{bmatrix}$	C-3
C-4)	$[+\text{comp}]_1^1 \rightarrow [+\text{long}]_1^1$ a) in the env		$+\begin{bmatrix}+\text{voc}\\-\text{cons}\end{bmatrix}$	C-4 a)
	b) in the env		$\emptyset$	C-4 b)
C-5)	$[\alpha\text{comp}] \rightarrow [\sim\alpha\text{comp}]$ in the env		$\emptyset$	C-5

All rules apply in the env $\begin{bmatrix}\overline{+\text{voc}}\\-\text{cons}\end{bmatrix}$, in extended strong stem only, i.e., the pre-final environment for rules C-1), C-2) is actually $\#C_o^3 \longleftrightarrow [+\text{son}]_1 \longleftrightarrow C_1$ +ext; for rule C-3): $\#C_o^3\,[\underline{\quad}]_1^n \begin{bmatrix}+\text{voc}\\-\text{cons}\end{bmatrix} C_1$ +ext; and for rules C-4), C-5): $\#C_o^3\,[\underline{\quad}]\,C_1$ +ext.

Low-level rules

p) Past $\rightarrow \emptyset$

q) VV $\rightarrow \overline{\text{V}}$

r) Diphthong structure: $[+\text{comp}]_3 \rightarrow [-\text{grave}]_2[+\text{grave}]$
(may be ordered before C-1)

Morphophonemically, long vowels are considered sequences of short vowels.

Class	Infinitive	Preterite Sg.	Preterite Pl.
1	rīdan	rād	ridon
2	bēodan	bǣad	budon
3	bindan	band	bundon
4	beran	bær	bǣron
5	metan	mæt	mǣton
6	faran	fōr	fōron

	-grave	+grave
+diffuse	i	u
{-diffuse, -compact}	e	o
+ compact	æ	a

OE Examples

(((# riid + ext) + Ø) + #)	Preterite Sg. – Class 1
(# ruud + ext)	C-1
(# raad + ext)	C-5
(# raad + ext + Ø)	New Cycle
(# rææd + ext + Ø + #)	C-1
(# rææd + ext + Ø + #)	New Cycle
(# raad + ext + Ø + #)	C-1
# rād #	Low-Level Rules
(((# bǽod + ext) + on) + #)	Preterite Pl. – Class 2
(# booed + ext)	C-1
(# baaæd + ext)	C-5
(# baaæd + ext + on)	New Cycle
(# bææad + ext + on)	C-1
(# biiud + ext + on)	C-2
(# bud + ext + on + #)	New Cycle
# budon #	Low-Level Rules
(((# bind + ext) + on) + #)	Preterite Pl. – Class 3
(# bund + ext)	C-1
(# band + ext)	C-5
(+ band + ext + on)	New Cycle
(# bænd + ext + on)	C-1
(# bind + ext + on)	C-2
(# bind + ext + on + #)	New Cycle
(# bund + ext + on + #)	C-1
# bundon #	Low-Level Rules
(((# met + ext) + on) + #)	Preterite Pl. – Class 4
(# mæt + ext)	C-5
(# mæt + ext + on)	New Cycle
(# mēt + ext + on)	C-4
(# mēt + ext + on + #)	New Cycle

OE Examples

(((# met + ext) + on) + #)	Preterite Pl. — Class 4
# mǣton #	Low-Level Rules
(((# far + ext) + Ø) + #)	Preterite Sg. — Class 5
(# fār + ext)	C-4
(# fōr + ext)	C-5
(# fōr + ext + Ø)	New Cycle
(# fōr + ext + Ø + #)	New Cycle
# fōr #	Low-Level Rules

T. G. Bever

References

1. S. J. Keyser, The Old English Weak Verb Cycle, Paper presented at Linguistics Seminar, Research Laboratory of Electronics, M.I.T., February 1963.

2. J. Wright and E. Wright, Old English Grammar (Oxford University Press, London, 1914).

3. E. Prokosch, A Comparative Germanic Grammar (University of Pennsylvania, Philadelphia, 1939).

G. THE E/O ABLAUT IN GREEK

The well-known alternations between the vowels e and o in verbal roots in Greek can be predicted from the constituent structure of the words in which these roots appear and a rule of the form

$$\text{A.}\quad [\alpha\text{grave}] \rightarrow [\sim\alpha\text{grave}] \text{ in env } \begin{bmatrix} -\text{diffuse} \\ -\text{compact} \\ -\text{cons} \end{bmatrix} \text{ X + Y (+\#) where X, Y do not contain \#}$$

operating in a transformational cycle.

To show how this rule operates, we require a statement of the internal constituent structure of Greek verbs and of nominals derived from verbs. Despite all of the attention that linguists have paid to these forms in the past, no such formulation has ever been attempted for Greek. Consequently, the formulation that I present in this report below must be regarded as being highly provisional. To substantiate the claims that I make, or to refute them, we require a thorough statement of Greek syntax so that we can see how the rules embodying these claims fit into such a syntax.

1. Constituent Structure of the Verb

Ignoring for the moment the low-level selectional restrictions, the phrase structure of the Greek verb may be schematized by means of the following phrase structure rules:

P1. $V \rightarrow$ (Augment) V_{base} Person Number (Voice)

P2. $V_{base} \rightarrow V_{stem}$ Theme (Mood)

P3. $V_{stem} \rightarrow V_{root}$ (Tense)

P4. Voice $\rightarrow$ Middle

P5. Mood $\rightarrow$ { Imperative / Optative / Subjunctive }

P6. Tense $\rightarrow$ { Present / Future / Aorist / Perfect / Passive (Future) / Aorist-Passive }

When the voice constituent is missing, the verb is said to be in the active voice; without the mood constituent it is in the indicative mood, and without the tense constituent, it is in the second aorist tense and the augment constituent must be present. Only certain verbs can appear in the second aorist. In a complete Greek grammar, we shall probably want to introduce the constituents person, number, voice, and mood transformationally, but the statement given here is adequate for our purposes. We also ignore the problems of how to handle the "primary" and "secondary" person endings of the active voice, and the person endings in the perfect tenses.

The theme constituent is missing in certain tenses of certain verbs: for example, in the present tense of verbs whose present tense constituent is nu. This constituent is also missing in the perfect middle of all verbs; its absence here can be handled by the following deletion transformation:

T1. Perfect Theme X Middle
1 2 3 4 $\rightarrow$ 1 0 3 4

The perfect constituent then permutes with the verbal root:

T2. V_{root} Perfect
1 2 $\rightarrow$ 2 1

An additional constituent is then inserted after the root in the perfect active:

T3. Perfect V_{root} Theme
1 2 3 $\rightarrow$ 1 2+K 3

The constituents K and perfect are rewritten in the morphophonemics by the rules:

M1. $K \rightarrow$ { k in env $[-\text{cons}])V_{root}$ + ——— / ϕ }

$$\text{M2. Perfect} \rightarrow \begin{cases} \text{a) e} \quad \text{in env} \text{ ---} + \begin{bmatrix} +\text{cons} \\ \alpha\text{vocalic} \end{bmatrix}_2 \\ \text{b) Ce} \quad \text{in env} \text{ ---} + C_1^1 \\ \text{c) VCV} \quad \text{in env} \text{ ---} + \begin{bmatrix} -\text{cons} \\ -\text{diffuse} \end{bmatrix} C_1^1 \\ \text{d) V} \quad \text{in env} \text{ ---} + \text{V} \end{cases}$$

Note that rule M2-c gives the "Attic reduplication."

Given this apparatus, we are able to predict the vocalism of verbal roots in e: we expect e-vocalism throughout the conjugation except in the perfect active tenses of verb roots that end in one or more consonants. Following the nomenclature of traditional Germanic grammars, we shall call such roots strong; roots that end in vowels, we shall call weak. Thus consider the conjugation of the verb from the strong verbal root streph, 'twist, turn': (we conjugate in the first person plural throughout)

1. Present active (#(((+streph + Ø)+e)+men)+#): stréphomen
2. Future active (#(((+streph + s)+e)+men)+#): strépsomen
3. Perfect active (#(((+e+streph)+e)+a+men)+#): estróphamen
4. Pluperfect active (#(+e((+e+streph)+e)+e+men)+#): ēstróphemen
5. Perfect middle (#(+e+streph)+metha)+#): estrámmetha

The a-vocalism in the perfect middle is a consequence of the zero-grade or vowel deletion rule, which I have not formulated here. Where we find full-grade in perfect middles of strong roots in e, we find the e-vocalism; thus for the root leip, 'leave', we have

6. Perfect middle (#(+le+leip)+metha)+#): leleímmetha

The conjugation of the weak verbal root kheu, 'pour' is exactly parallel to that of streph, except that in the perfect active tenses, the morpheme k given by rule M1 is present and the e-vocalism is maintained:

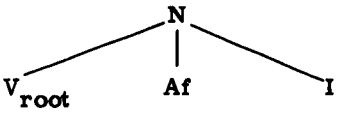
7. Perfect active (#(((+khe+kheu+k)+e)+a+men)+#): kekheúkamen.

Since the only form in which rule A applies an odd number of times to the vowel of the verbal root is in the perfect active tenses of strong verbs, only there do we find the o-vocalism in the root. In every other form it operates an even number of times; twice in the perfect middle and four times elsewhere.

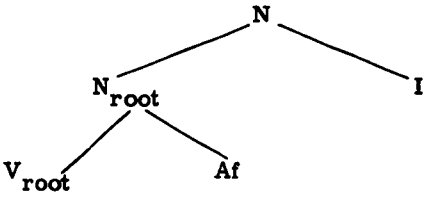
2. Constituent Structure of Nominals Derived from Verbal Roots

It seems reasonable to suppose that the derivation of nominals from verbs in Greek is merely a special case of generalized transformations that embed into noun or adjective phrases deformations of full sentences in which the verb appears. The verbal root, together with a derivational affix, becomes the head noun or adjective in the phrase. The resulting noun must obtain a grammatical gender, and presumably it gets it from

derivational affix. I claim, for the moment without justification, that neuter and common gender affixes behave differently from affixes that are inherently feminine or masculine (by common gender affix, I mean an affix that is not inherently either masculine or feminine, but, depending on the derived word, it takes on one or another of these genders). The derived constituent structure of nouns having the first type of derivational affix is



where I is the inflectional suffix, supplied transformationally, which marks the case and number of the noun. Nouns formed with the second kind of derivational affix have the structure



The neuter derivational affixes are es and mat: the common gender affixes are the agentive affixes tēr and tor; inherently feminine affixes are ā, id, and ad; inherently masculine affixes are o and eu. Given this derived structure, we are immediately able to account for the vocalism in derived nouns from verbal roots in e: derived neuter and agent nouns will have e-vocalism, whereas derived feminine and masculine nouns formed from the affixes just listed will have o-vocalism as illustrated by the following examples, in which we use the roots streph and blep, 'look':

8.	(#(+streph+mat+os)+#)	:	strémmatos, 'a twist'	Genitive singular
9.	(#(+blep+es+os)+#)	:	blépeos, 'a look'	
10.	(#(+streph+ter+os)+#)	:	streptêros, 'anything which turns'	
11.	(#((+streph+o)+s)+#)	:	stróphos, 'a band'	Nominative singular
12.	(#((+streph+eu)+s)+#)	:	stropheús, 'a vertebra'	
13.	(#((+streph+ā)+Ø)+#)	:	strophā́, 'a turning around'	
14.	(#((+streph+id)+s)+#)	:	strophís, 'a girdle'	
15.	(#((+streph+ad)+s)+#)	:	strophás, 'a circling'	

Weak verb roots that end in diffuse vowels also form derived nouns with the same ablaut relationship. For example, from the root kheu

16.	Genitive singular	(#(+kheu+mat+os)+#)	:	kheúmatos, 'a stream'
17.	Nominative singular	(#(+kheu+o)+s)+#)	:	khóos, 'a liquid measure'

18. Nominative singular (#(kheu+ā)+Ø)+#) : khoá, 'a libation'
and from the root dei, 'fear'
19. Genitive singular (#(+dei+es+os)+#) : déeos, 'fear'

3. Extent of Ablaut in Greek

Verbal roots with fundamental vocalism o rather than e do not undergo ablaut. Thus, although the verbal root kop 'strike' is strong, the vocalism of the perfect is not different from that of the other tenses; the present active is kóptomen, and the perfect active is kekóphamen. Similarly, all nouns derived from this root have an o-vocalism: for example, kómmatos, genitive singular, 'that which is struck'; kópos, nominative singular, 'a striking'; and kopís, nominative singular, 'a chopper'. Furthermore, all pure noun roots never exhibit ablaut. Certain derivational and inflectional endings, however, do show it, and in those cases in which it appears, it can be handled by rule A as before. For example, the neuter affix es appears as os in the nominative and accusative singular. But since there is a rule in Greek which deletes the nominative and accusative singular marker in all neuter nominals not formed with the affix o, this alternation follows immediately from rule A. Consider the nominative and accusative singular of forms 9 and 19:

20. (#(+blep+es)+#) : blépos
21. (#(+dei+es)+#) : déos

The masculine affix o of examples 11 and 17 appears as e in the vocative singular because the vocative singular marker has been deleted, so that rule A applies one less time to it. The affix eu, however, does not change in the vocative singular.

The theme of the verbal conjugation, which we have written e, also undergoes ablaut, but rule A alone cannot give the correct results. If, however, we suppose that there is also a low-level phonetic rule in Greek,

$$e \rightarrow o \quad \text{in env} \_\_\_\_ + [+\text{nasal}],$$

then the alternations of the theme vowel can be handled.

It is convenient, then, to set up a class of ablauting vowels in Greek: the e of verbal roots, and the vowels of certain inflectional morphemes. We are not forced, however, to mark the distinction between ablauting and nonablauting vowels phonemically; we require only a morphophonemic rule that will specially mark the ablauting vowels, and an adjustment to rule A which will allow it to operate only on those vowels that are so marked.

I am at present working on a system of rules which will account for the loss of vowels in certain positions (the zero-grade phenomenon), vowel lengthening, and accent placement. This entire system is also, apparently, part of the transformational cycle in Greek.

D. T. Langendoen

XV. COMMUNICATIONS BIOPHYSICS*

Prof. W. A. Rosenblith
Prof. M. A. B. Brazier†
Prof. M. Eden
Prof. M. H. Goldstein, Jr.
Prof. W. T. Peake
Prof. W. M. Siebert
Dr. J. S. Barlow‡
Dr. A. Cavaggioni**
W. A. Clark††
Dr. B. G. Farley††
Dr. G. L. Gerstein
Dr. E. Giberman‡‡
Dr. R. D. Hall
Dr. N. Y-S. Kiang***
Dr. T. T. Sandel††
Dr. Eda Berger Vidale
J. A. Aldrich
R. M. Brown
S. K. Burns
R. R. Capranica
Eleanor K. Chance
R. J. Clayton
A. H. Crist
P. R. Gray
J. L. Hall II
F. T. Hambrecht
J. G. Krishnayya
R. G. Mark
P. Mermelstein
C. E. Molnar†††
Donna A. Molnar
R. R. Pfeiffer
Cynthia M. Pyle
D. M. Snodderly, Jr.
G. F. Svihula
Aurice V. Weiss
T. F. Weiss
J. R. Welch
M. L. Wiederhold
G. R. Wilde

A. DESIGN PHILOSOPHY FOR PSYCHOPHYSICAL EXPERIMENTS RELATED TO A THEORY OF AUDITORY FUNCTION

Psychophysics should play a major role in the evolution of a theory of auditory function, together with anatomy and physiology. We feel that the psychophysical tests that will be most valuable for this purpose are those that can be related to the physiological information now available about hearing. In terms of this objective of relating psychophysical and physiological data, it is possible to argue for a "best" psychophysical test procedure.

A good deal is known about the mechanical properties of the ear,[1] and about the neural activity at the level of the acoustic nerve,[2] that is, about what might be called the peripheral auditory system. In particular, it is now clear that some information about the external acoustical stimulus is destroyed by the peripheral auditory system, and that psychophysical test performance is thereby fundamentally limited. By utilizing the available physiological information about the periphery in a statistical hypothesis testing formulation, the limits on performance imposed by the periphery in certain psychophysical tests could be described quantitatively, except, perhaps, for a few unknown parameters and functional forms. The more central system also imposes

---

*This work was supported in part by the National Science Foundation (Grant G-16526); and in part by the National Institutes of Health (Grant MH-04737-02).

†Visiting Professor in Communication Sciences from the Brain Research Institute, University of California at Los Angeles.

‡Research Associate in Communication Sciences from the Neurophysiological Laboratory of the Neurology Service of the Massachusetts General Hospital.

**From Istituto di Fisiologia, Università di Pisa.

††Staff Member, Lincoln Laboratory, M.I.T.

‡‡From the Department of Physics, Weizmann Institute of Science, Israel.

***Also at the Massachusetts Eye and Ear Infirmary, Boston.

†††Air Force Cambridge Research Laboratories.

limits on performance, but these more central limitations cannot be described quantitatively on the basis of physiological information, as present understanding of what is done centrally with the information available from the periphery is relatively incomplete. It may be that in some psychophysical tests, the dominant limitations on performance are peripheral; that is, the central processing is effectively as good as the mathematically optimal processing. If such is the case, the quantitative description of peripheral limitations for these tests should correctly describe the characteristics of the psychophysical data. Since some psychophysical data must be used to complete the formulation of peripheral bounds, the only fact that can really be demonstrated experimentally is that there exists a class of psychophysical tests in which the data are as if peripheral bounds were dominant. We can look for a class of psychophysical tests in which the data from one test are related to the data from another in the same way as they would be if peripheral limitations were dominant, but, even if we are successful, we cannot claim to have demonstrated that peripheral limitations are really dominant. However, this approach seems to be a legitimate and promising way in which to try to relate physiological and psychophysical phenomena.

If we subscribe to this approach, we must look first for psychophysical experiments that can be related to the statistical hypothesis testing formulation of peripheral bounds and hold some promise of being tests in which peripheral factors are dominant. Given these conditions, we can argue for several desirable attributes of psychophysical tests:

(i) The tests should involve discriminations rather than absolute recognitions.

(ii) The subject should be thoroughly informed about every aspect of the experimental design except the actual sequence of presentations to be used (the subjects used in the experiments described here were graduate students in electrical engineering, and hence had no difficulty in understanding a quantitative description of the stimuli).

(iii) The subject should be asked to choose from a finite and preferably small set of responses, say, two.

(iv) The stimulus parameters should unambiguously define which response is correct and the subject should be informed of which response was correct after each trial.

(v) The subject should be asked to use all of the information available to him to do as well as possible in terms of some scoring system.

(vi) There should be a minimum of different possible stimuli, say, two.

These points lead to consideration of two general types of discrimination tests, each of which involves only two possible stimuli and two possible responses. In one configuration, the subject is presented stimulus A followed by stimulus A (AA), or stimulus A followed by stimulus B (AB), and asked whether the two stimuli are the same or different. In the other configuration the subject is presented either AB or BA and asked in which position is B.

A series of experiments in frequency discrimination was run to evaluate the relative

merits of variations of these two presentations. Stability of performance was a major consideration in evaluating the various psychophysical test schemes, since unstable behavior was regarded as a clear indication that peripheral bounds were not dominant. We found that the symmetric presentations – those in which either pair occurred with a priori probability 1/2 – were better in terms of stability than the asymmetric. This conclusion was supported by the subjective feelings of the observers and the quantitative aspects of the experimental data. We also found that the High-Low (AB, BA) presentation was better in terms of stability than the Same-Different (AA, AB). The subjects preferred the (AB, BA) configuration because they felt that high-low was exactly the question they first answered in listening to the stimuli, whereas in the (AA, AB) configuration the stimuli all sounded different and they asked themselves "different enough?" The subjects' feelings were reflected quantitatively in the data, with the (AB, BA) data being significantly more stable than the (AA, AB). We feel that the two-alternative forced-choice experiment (AB, BA) would also be "best," given the goal of relating physiological to psychophysical data, in tests of discrimination other than frequency and in detection tests.

Further details of these arguments, the statistical hypothesis testing formulation, and the experiments referred to in this report may be found in the author's thesis.[3]

P. R. Gray

References

1. G. von Békésy's papers have been collected in Experiments in Hearing (McGraw-Hill Book Company, Inc., New York, 1960).

2. H. Davis, Advances in the neurophysiology and neuroanatomy of the cochlea, J. Acoust. Soc. Am. 34, 1377-1385 (1962).

3. P. R. Gray, A Design Philosophy for Psychophysical Experiments, S. M. Thesis, Department of Electrical Engineering, M.I.T., January 1963.

B. EVOKED RESPONSES IN RELATION TO VISUAL PERCEPTION AND OCULOMOTOR REACTION TIMES IN MAN

[This report is a summary of an invited paper read at the Conference, "Sensory Evoked Response in Man," sponsored by the New York Academy of Sciences and Albert Einstein College of Medicine, New York, February 14-15, 1963. The paper will be published as a part of a Conference Monograph in the Annals of the New York Academy of Sciences.]

Responses evoked at the occiput in man by visual stimuli have been studied in relation to oculomotor reaction times, and in relation to some of the associated phenomena of visual perception. Sudden changes in the vertical position of a spot on an oscilloscope screen were used as stimuli and as tracking signals, and motor responses consisted of the corresponding redirections of gaze to the successively new positions of the spot.

For subsequent analysis, simultaneous recordings were made onto magnetic tape and paper of: (a) spot position, (b) eye position (monitored electro-oculographically), and (c) EEG potentials recorded bipolarly from a pair of midline parieto-occipital scalp electrodes.

Averages of parieto-occipital evoked potentials for a large number of stimuli were obtained from recordings made during visual tracking of the spot, and also from recordings carried out while gaze was maintained fixed at the center of the oscilloscope screen. Also, averaged responses, with eyes closed, to flashes of light from a stroboscope were obtained. Oculomotor reaction times were determined by averaging of electro-oculographic potentials, and by crosscorrelation of the latter with the spot-position signals.

In separate psychophysical experiments with some subjects, an attempt was made to determine, by indirect means, the approximate time interval for subjective perception of the stimuli, and for subjective perception of eye movements, for comparison with the latencies of various components of the parieto-occipital evoked responses.

Potentials evoked by the shifting spot of light were generally appreciably smaller, and greater in latency, than those evoked by bright flashes of light with the eyes closed, but the amplitude of certain components of the former (that is, those with an output at approximately 80-90 msec) was clearly increased during tracking for some subjects, an effect that may have been due to the fact that the fovea of the retina was more frequently stimulated during tracking than it was when the eyes were maintained fixed. Consideration of the latencies of these components suggests that they may represent nonspecific rather than specific evoked responses in the visual system. Comparison of averages of EEG responses with those of EOG responses established that the above-mentioned components appearing during tracking did not represent the electrical field, at the parieto-occipital scalp electrodes, of the EOG potentials, the amplitude of the latter being much larger than that of the EEG potentials. Suggestive, but not conclusive, evidence for different forms of evoked responses for different directions of spot-shifts were obtained for some subjects.

A comparison, for one subject, of the results of the psychophysical experiments with those of the electrophysiological recordings provided some tentative evidence that the subjective perception of spot-shifts takes place, on the average, rather early (that is, within 50 msec), a finding that suggests that visual perception of simple stimuli may be more closely related to specific visual pathways than to nonspecific ones. Only minimal evidence of an early, specific response was, however, obtained in the present series of experiments for spot-shifts as visual stimuli. It appears probable that a component appearing in the averaged parieto-occipital EEG response at a constant interval of time after the mean oculomotor reaction time represents the same basic phenomenon as that of the "lambda waves" following eye movements

that have been described previously.

The findings from these electrophysiological and psychophysical experiments are discussed in relation to the question of quantization of time in the nervous system.

This work was supported in part at the Massachusetts General Hospital by a U. S. Public Health Service career program award (Number 5-K3-NB-9201), and by a Public Health Service research grant (Number B-3752), from the National Institute of Neurological Diseases and Blindness.

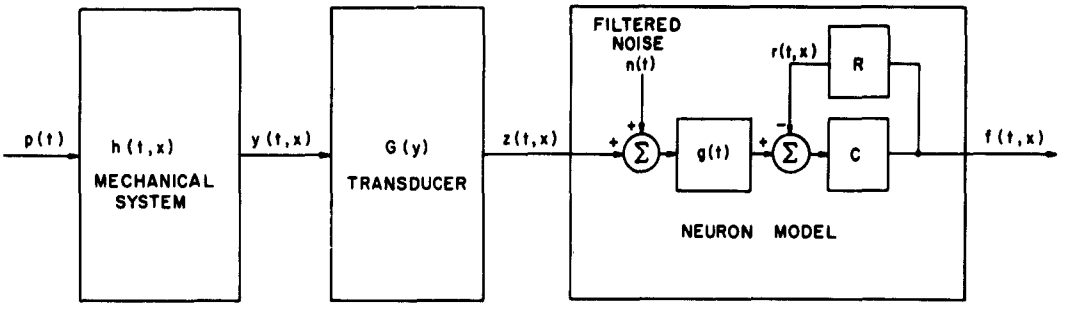
J. S. Barlow

C. A MODEL FOR FIRING PATTERNS OF AUDITORY NERVE FIBERS

Acoustic signals are represented by sequences of spike potentials in the auditory (VIIIth) nerve. The form of this code and the encoding mechanism have been subjects of considerable interest in physiology. Recent experimental results of Kiang and others[1] have revealed the highly systematic structure of the spike activity of fibers in the VIIIth nerve of cats. These results have led to the work, summarized here, on the construction and testing of a model of the peripheral auditory system that relates the firing of an VIIIth-nerve fiber to sound coming into the ear.

Sound enters the outer ear, impinges on the eardrum, and is transmitted through the middle-ear structures to the fluids of the cochlea or inner ear. The fluid motion results in motion of the cochlear partition, which in turn results in forces on and/or movements of the auditory receptor cells or hair cells. The hair cells are thought to be transducer elements whose function is to produce local excitation of nearby VIIIth-nerve fibers. The spike potentials resulting from this excitation are transmitted to the brain via the VIIIth nerve.

Figure XV-1 shows a model of this system. The "Mechanical System" represents the functional relationship between an acoustic pressure input to the ear and a displacement of the cochlear partition and is assumed to be representable as a linear system over a range of approximately 80 db of sound intensity. The transfer function of this part of the system for a particular point along the cochlear partition is assumed to be given by the work of von Békésy[2] for frequencies below approximately 2 kc. The "Transducer" is intended to represent the functional relationship between the displacement of the cochlear partition at a point along its length and the output of a hair cell at that point. The final block shows a stereotyped or idealized "Model Neuron." In this model the output of the transducer is summed with noise at the input of the neuron. This noise is included to account for both the spontaneous activity and the probabilistic response behavior characteristic of VIIIth-nerve fibers. The existence of this noise (additive to the membrane potential) can be justified to some extent on the basis of extrapolations from some empirical evidence that is due to Verveen.[3] The sum of the



p(t), Pressure at the ear drum

y(t,x), Displacement of the cochlear partition

z(t,x), Output of hair cell

f(t,x), Action potentials of fiber

h(t,x), Impulse response of the cochlear partition at a point x cm from the stapes

G(y), A function relating the output of a hair cell to displacement of the cochlear partition at a point x cm from the stapes

g(t), Impulse response of a linear filter

r(t,x), Threshold of the neuron model at time t

Fig. XV-1. Model relating the firing patterns of fibers in the auditory nerve to acoustic stimuli.

noise and the input to the neuron is filtered and then compared with a threshold in the box labelled "C." If the threshold is exceeded, then a spike is defined as occurring, and the threshold is reset to some larger value, by the box labelled "R."

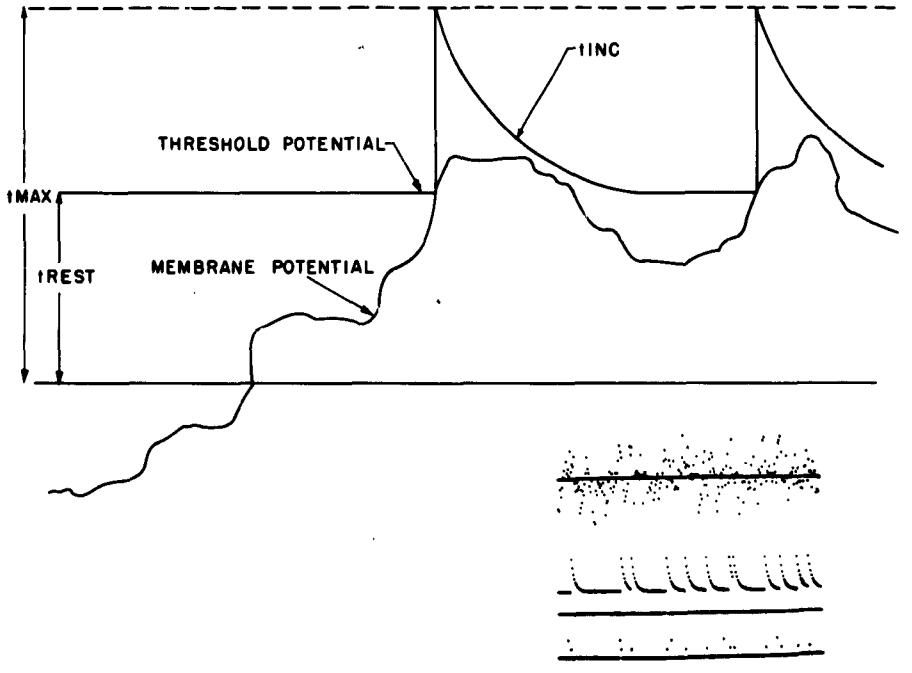


Fig. XV-2. Threshold and membrane potential of the model neuron as a function of time.

Figure XV-2 shows both the noisy membrane potential of the model neuron and the threshold as a function of time. The threshold is reset to some large value (tMax) upon the occurrence of a spike and decays to its resting value (tRest) with a decrement (tInc). In the lower right-hand corner there is a representation of the membrane and threshold potentials generated by the TX-2 computer simulation of this system. The top trace shows the noisy membrane potential. The center trace shows the threshold as a function of time. The bottom trace of dots shows the times of occurrence of spikes and a set of time markers.

The fundamental assumptions in this model are:

(i) The mechanical system is assumed to be representable by a linear system.

(ii) A point-to-point relation between the displacement of the cochlear partition and the neural excitation is assumed. A particular neural fiber is assumed to be excited by a neighboring hair cell, which in turn responds to the displacement of the cochlear partition at a point along its length.

(iii) The process of neural excitation is represented by a simple model neuron. This model is probabilistic and characterized by threshold and refractory effects.

(iv) The effects of efferent fibers on the peripheral system are ignored.

This model has been simulated on the TX-2 computer at Lincoln Laboratory, M.I.T.,

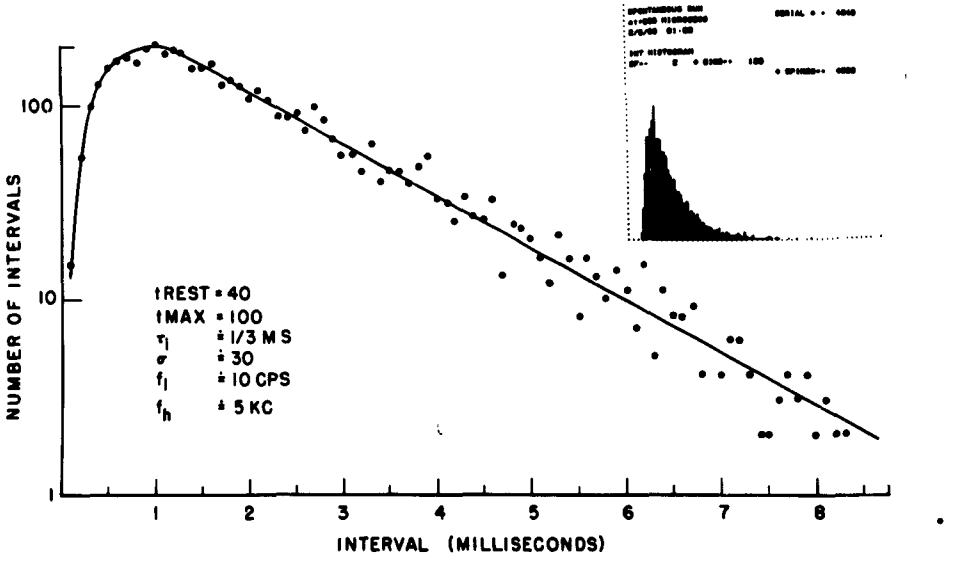


Fig. XV-3. Histogram of intervals between spikes generated by the model in the absence of a stimulus. The time constant of the decay of the threshold to its resting value is 1/3 msec. The noise spectrum is flat from 10 cps to 5 kc.

Fig. XV-4.

Impulse response of the cochlear partition at a point whose characteristic frequency is 500 cps. Dot separation is 100 μsec, and successive maxima of the response are separated by 2 msec. This particular impulse response corresponds to $F_3(t)$ (see Flanagan[4]).

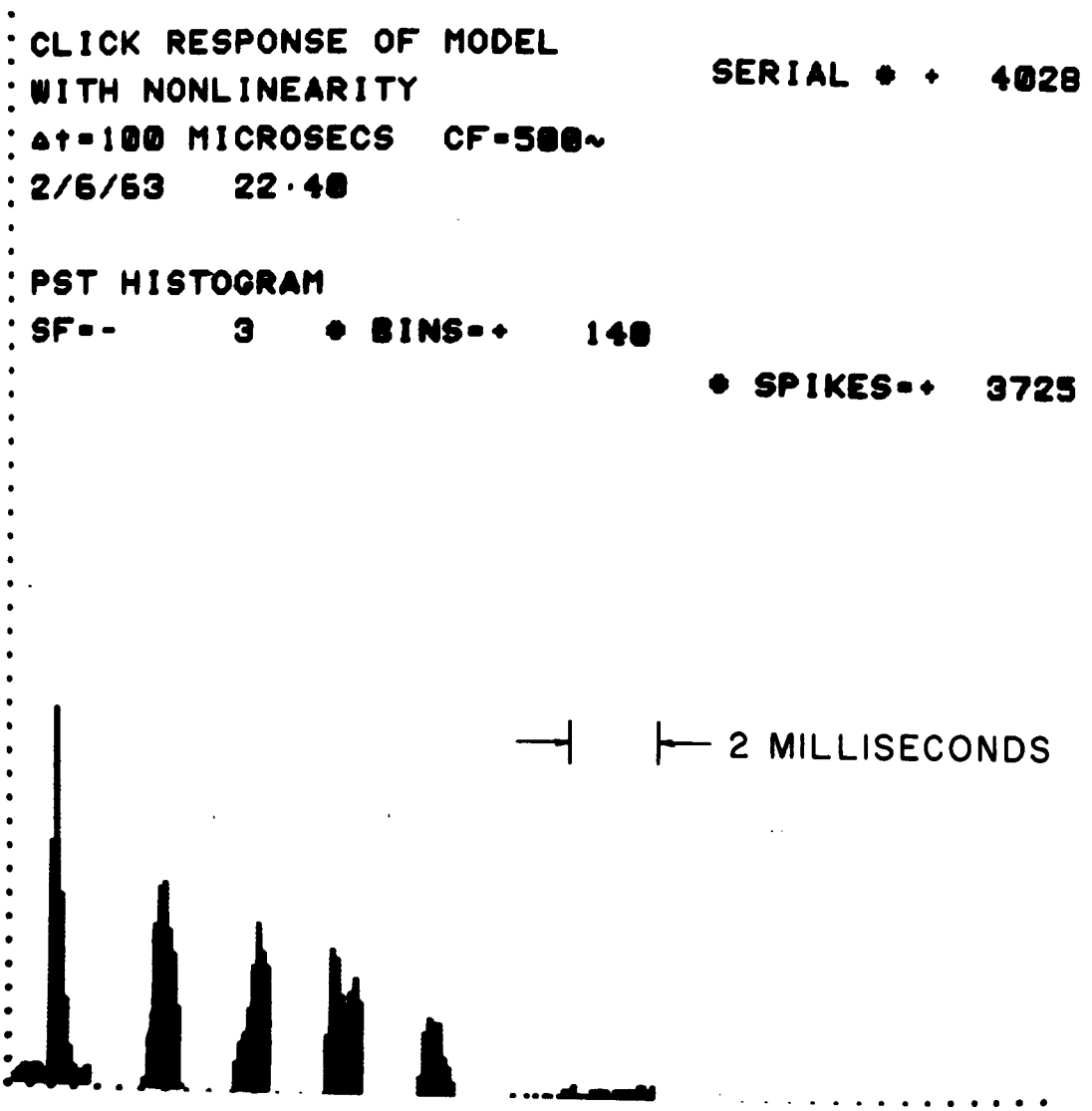


Fig. XV-5. PST histogram of the click response of the model for a fiber whose characteristic frequency is 500 cps.

and statistics of the spontaneous activity and response of the model to a variety of acoustic stimuli have been studied and compared with the VIIIth-nerve data obtained by Kiang and others.[1]

Figure XV-3 shows a histogram of interspike intervals of the spontaneous activity generated by the model. The insert shows this histogram as photographed from the TX-2 oscilloscopic output. The rest of the figure shows the interval histogram plotted in semi-logarithmic coordinates. Notice that the tail of the distribution is linear with these coordinates, or exponential on linear coordinates. For very short intervals, the the number of firings decreases to zero, a reflection of the refractory properties of the model. This general form of the distribution of intervals in the spontaneous (or unstimulated) case is characteristic of VIIIth-nerve fibers.

Figure XV-4 shows the response of the cochlear partition to a short click determined by Flanagan[4] from the data of von Békésy. In this case, the response is given for a point along the cochlear partition whose characteristic frequency is 500 cps. Figure XV-5 shows the response of a model fiber that is assumed to innervate a hair cell at the same spot on the cochlear partition. The response is in the form of a PST (Post Stimulus Time) histogram. This histogram shows the number of firings that occur at a time t after the onset of a stimulus. The peaks in this histogram are separated by an interval of 2 msec and correspond to the times of the maximum positive deflections in the response of the cochlear partition at the 500-cps point to the click. This general form of the model PST histogram agrees quite well with the form of similar histograms computed from the VIIIth-nerve data.

Figure XV-6 shows PST histograms of the click response of the model for two different forms of the transducer as a function of sound intensity. The histograms on the left side show the results for a linear transducer, that is, the output of the transducer is proportional to its input. Note that there is a 36-db range between the intensity at which a response first appears to be visible in the PST histogram and the intensity at which the response of the model becomes stereotyped (that is, the model responds to the largest positive deflection of the cochlear partition every time a stimulus is presented.) This behavior does not agree with the observed response of VIIIth-nerve fibers. Since the envelope of the impulse response of the cochlear partition decays at a rate exceeding 20 db per cycle of oscillation, there can be at most two or three peaks in the PST histogram when a linear transducer model is used. The linear transducer function thus yields PST histograms that are at variance with the empirical evidence of Kiang and others. The histograms on the right side of Fig. XV-6 show the results obtained for a transducer model that is a nonlinear saturation function. The results obtained with the nonlinear transducer model agree qualitatively with the results of Kiang and others.

One of the puzzling results of the investigations of Kiang and others indicates that

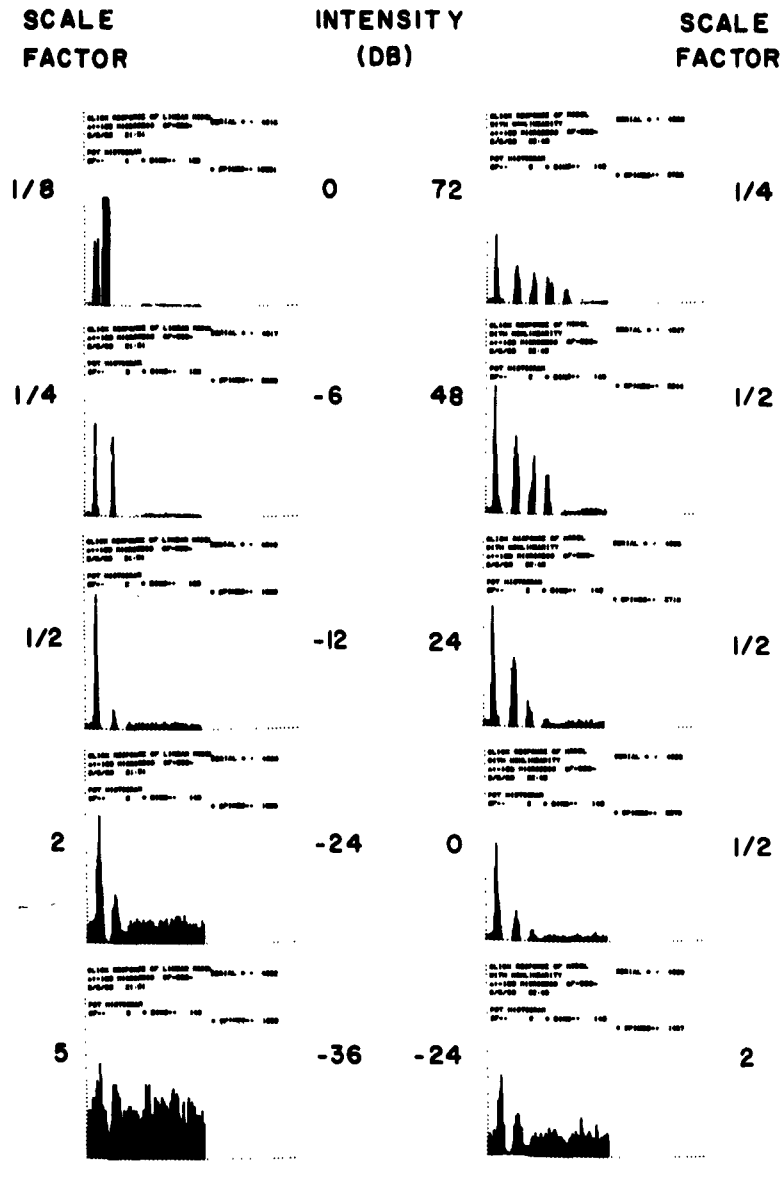


Fig. XV-6. PST histograms of the click response of the model for two different transducer functions vs intensity of click. The left-column histograms were computed for a linear transducer. The right-column histograms were computed for a nonlinear transducer $\left[G(y) = y \frac{k}{k + |y|}\right]$.

the sensitivity of neurons responding to sinusoids at their characteristic frequency increases as a function of frequency (at a rate of approximately 40 db/decade) in the range 100 cps-2 kc. Such a change in sensitivity is seen at higher auditory centers[5] and even in the human audiogram,[6] but it was surprising to encounter this effect so close to the periphery, particularly since the mechanical part of the system does not exhibit so large a change in sensitivity. Preliminary results tend to indicate that this effect can be explained by the model. That is, the model predicts a change in sensitivity as a function of frequency which is qualitatively the same as the change in sensitivity seen in VIIIth-nerve fibers, but the quantitative results depend upon the definition of the threshold of firing of a unit and a knowledge of the exact parameters of the system.

To summarize, this model of the peripheral auditory system seems capable of fitting certain available VIIIth-nerve data over a considerable range of stimuli. The important constituents of the model are a linear bandpass filter, followed by a nonlinear saturation function, followed by a probabilistic threshold device with refractory properties.

T. F. Weiss

References

1. N. Y-S. Kiang, T. Watanabe, Eleanor C. Thomas, and Louise F. Clark, Stimulus coding in the cat's auditory nerve, Ann. Otol. Rhinol. Laryngol. 71, 1009-1016 (1962).

2. G. von Békésy, Experiments in Hearing, edited by E. G. Wever (McGraw-Hill Book Company, Inc., New York, 1960).

3. A. A. Verveen, Fluctuation in Excitability (Drukkerij Holland N. V., Amsterdam, 1961); Axon diameter and fluctuation in excitability (in press).

4. J. L. Flanagan, Models for approximating basilar membrane displacement, Bell System Tech. J. 39, 1163-1192 (1960).

5. J. E. Hind, Unit activity in the auditory cortex, Neural Mechanisms of the Auditory and Vestibular Systems, edited by G. L. Rasmussen and W. F. Windle (Charles C. Thomas, Springfield, Ill., 1960), p. 203.

6. J. C. R. Licklider, Basic Correlates of the Auditory Stimulus, Handbook of Experimental Psychology, edited by S. S. Stevens (John Wiley and Sons, Inc., New York, 1951), p. 995.

D. CORTICAL FACILITATION FOLLOWING ONSET OR TERMINATION OF A LIGHT

The present report concerns the shock-evoked response complex (SERC) recorded from the primary visual cortex. This response has been the subject of considerable study. The complex of deflections is obtained whether one delivers the "test" shocks to the optic nerve, chiasm, tract, the lateral geniculate body or the optic radiation. (These deflections are usually numbered as illustrated in Fig. XV-8, trace D.[1,2]) There is general agreement that deflection no. 1 is related to the volley of impulses in nerve

fibers that feed into the cortex, and that deflections no. 3, no. 4, and no. 5 are measures of intracortical events resulting from this influx. The SERC for a given size of test shock may be modified by changing the animal's physiological state, or by changing conditions of retinal stimulation. The most striking modification of this SERC is an enhancement, which has been demonstrated under a number of conditions, including arousal by reticular stimulation or other means,[3,4] presentation of a continuous diffuse illumination,[5] and retinal deafferentation.[6] Besides these effects which last long, there are transient changes in the SERC following a sensory or arousing stimulus. In this report we shall deal with such a transient effect; specifically, with a strong enhancement following the onset or termination of the presentation of a light. The effect is most clearly demonstrated when the shocks are delivered to the optic radiation rather than to more peripheral loci in the afferent pathway, and it is abolished by deep barbiturate anesthetization. The experiments were performed with unanesthetized cats with mid-pontine pretrigeminal brain-stem sections.

Recording was by a monopolar electrode (silver ball or cotton wick soaked in NaCl solution), with the neutral electrode on the skin of the neck. Shocks were delivered by concentric electrodes. The light source was a Westinghouse 47 bulb placed in the end of a hollow cylinder, 18 cm long and 3 cm in diameter. At the other end of the cylinder was a glass diffuser that was placed directly in front of the cat's eyes at a distance of approximately 8 cm. The bulb was excited by a 6-volt dc source. The electrocortical

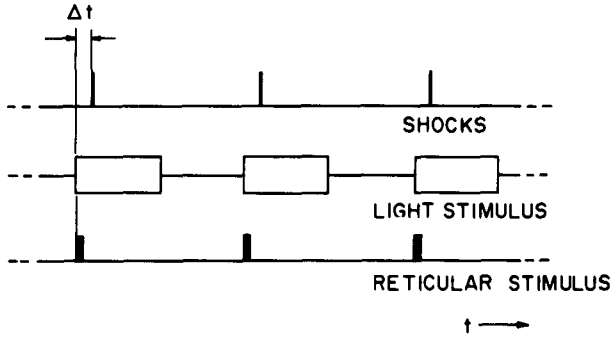


Fig. XV-7. Temporal pattern of stimuli. Top line shows the timing of the shocks to the optic pathway. Stimuli are delivered repetitively at the rate of 1 per 2 seconds or 1 per 4 seconds. In control conditions the cat is in the dark. When light or reticular stimulation is coupled with the shocks the time sequence is as shown, with $\Delta\tau$ the time between the onset of the electric signal to the light, or the train of shocks to the reticular formation (represented by a solid bar) and the shock to the optic pathway. (For the time course of the light onset see Fig. XV-8.) Duration of the light stimulus is 1 sec; duration of the train of shocks to the reticular formation, 0.1 sec.

activity was recorded on magnetic tape and processed by the ARC-1 average response computer.

Figure XV-7 illustrates the temporal pattern of stimuli used in the experiments. The average waveform of cortical responses to the onset of the light is shown in Fig. XV-8, trace A. Figure XV-8, trace B shows, on the same time and amplitude scale as in trace A, the average of responses to the light together with a single shock to the optic

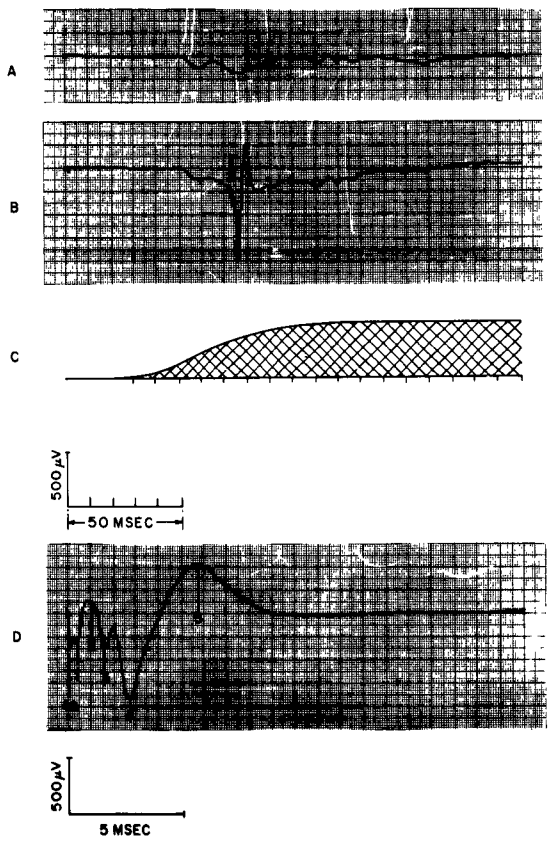


Fig. XV-8. Evoked responses from primary visual cortex. A: average of responses to the onset of the light. B: average of responses to the onset of the light together with a shock to the optic radiation homolateral to the recording electrode delivered at the time indicated by the arrow. C: time course of the light as measured by a photomultiplier tube. The trace starts at the onset of the electric signal to the bulb. D: detail of the part of the waveform in B following the shock. The shock artefact (s. a.) and characteristic deflections of the shock-evoked response complex (SERC) are indicated in this and subsequent figures. The shocks are 4 v in amplitude, and 0.2 msec in duration, and presented once every 4 seconds. In this and subsequent figures the number of responses in each average is 32.

radiation delivered at the time marked by the arrow. Figure XV-8, trace C shows the time course of the onset of the light. It is clear that the SERC is larger and of much shorter duration than the response to the light alone, so that large changes in the sizes of its deflections cannot result from an addition of these two responses. Figure XV-8, trace D shows a more detailed view of the SERC on a different time scale.

The enhancement effect is illustrated in Fig. XV-9. Figure XV-9, trace A shows the SERC (shocks delivered to the optic radiations) with control (dark) conditions. Figure XV-9, trace B shows the SERC when the shocks are delivered 150 msec following the electric signal to the light source. (Note that there is an approximate 30-msec delay between this event and the light onset, see Fig. XV-8, trace C.) It is interesting to

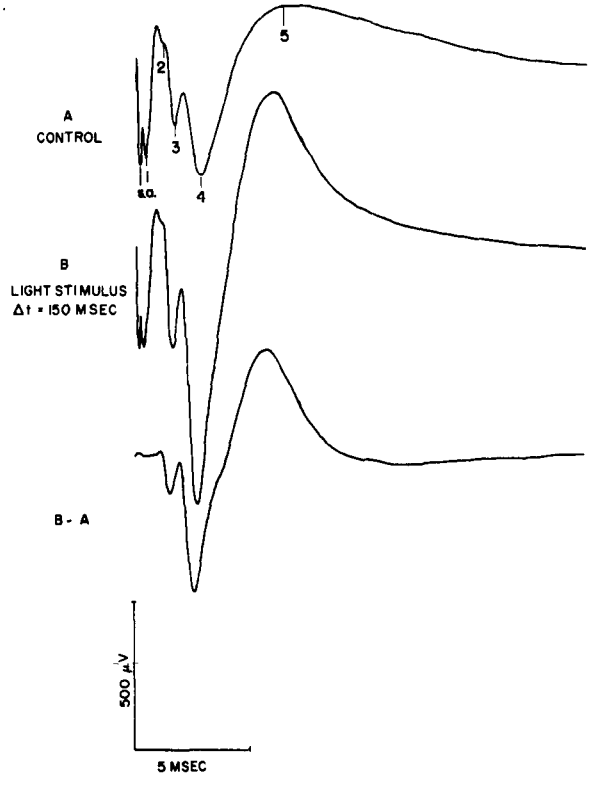


Fig. XV-9. Enhancement of the SERC by the onset of a light. A: average of responses to shocks to the optic radiation under control conditions of darkness. B: average of responses to shocks together with the presentation of light. The third trace is the difference between the waveforms obtained in B and A. The shocks were given 150 msec after the electric signal to the light. Other stimulus parameters are the same as in Fig. XV-8.

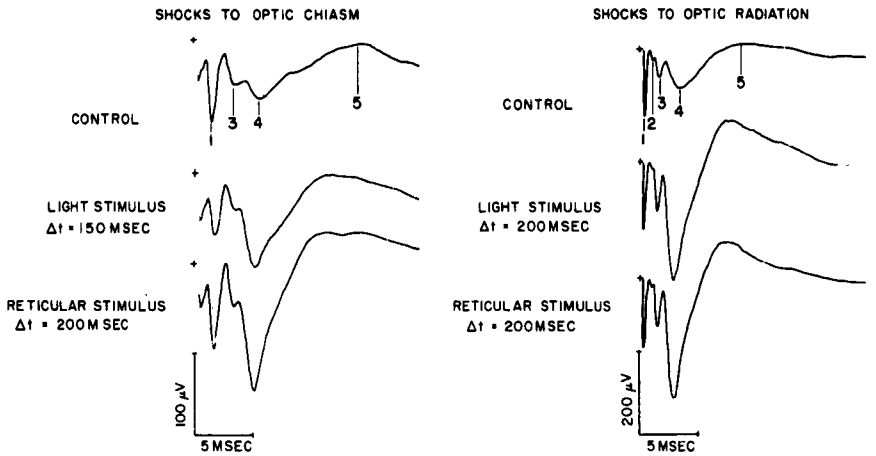


Fig. XV-10. Modification of the SERC by the onset of a light and by reticular stimulation. Left column shows averages of responses to shocks to the optic chiasm, and their modification by the onset of a light and by the onset of a train of shocks to the reticular formation. Right column gives similar results for shocks delivered to the optic radiation. The stimuli were delivered repetitively at a rate of 1 per 2 seconds. The shocks to the chiasm were 4 v in amplitude, 0.2 msec in duration; those to the optic radiation were 3 v in amplitude, 0.2 msec in duration. Reticular stimuli were 0.1-sec trains of shocks having 2-volt amplitude, 300/sec rate, and 0.2-msec duration.

observe that only deflections no. 3, no. 4, and no. 5 are modified significantly. The differential waveform of Fig. XV-9 illustrates this point. A similar enhancement effect may be demonstrated following termination of the light.

The rate of presentation of the light is important in determining the extent of enhancement. For rates of 1 per second, or greater, the enhancement is small. In several cases a considerably greater enhancement was observed for the rate of 1 per 4 seconds than for the rate of 1 per 2 seconds. The enhancement effect is obtained for a wide range of light intensities, and can be demonstrated with the stimulating light shining directly into the animal's eyes, or with the light overhead.

In a number of experiments we have attempted to enhance the SERC by reticular stimulation. In these midpontine pretrigeminal preparations such an enhancement is possible, but difficult to obtain. The effect is extremely sensitive to the location of the stimulating electrode in the reticular formation. When an enhancement is obtained it occurs in the period following the onset of the train of shocks to the reticular formation, and cannot be demonstrated during a continuous reticular stimulation. This enhancement is greatest if the test shocks are delivered to the optic pathway approximately 200 msec after the onset of

the train of shocks to the reticular formation. A 100-msec train of shocks to the reticular formation seems to provide as much enhancement as a longer one. Figure XV-10 shows enhancement of the SERC for shocks to the optic chiasm and enhancement of the SERC to shocks to the optic radiations following reticular stimulation (third waveform in each column). Note that in both cases the size of deflection no. 1 is practically unchanged, while deflections no. 3, no. 4, and no. 5 are enhanced.

Figure XV-10 also shows corresponding modification of the SERCs for shocks to the optic chiasm and optic radiations following the onset of a light (second waveform in each column). For shocks delivered to the optic radiations results are very similar to those obtained with reticular stimulation. However, for shocks delivered to the optic chiasm the modification is more complex; deflection no. 1 is smaller than in the control, while the later events show some enhancement.

We interpret these results as follows: Following the onset or termination of a light there is a facilitation of the neurons of the primary visual area of the cortex. This facilitation is demonstrated by the "test probe" of a shock to the optic radiations. There is an enhancement of the intracortical events 3, 4, and 5 in the SERC when the test shock is delivered in the appropriate time interval following the onset or termination of the light. The modification of the SERC when the test shock is delivered to the chiasm demonstrates an inhibitory effect which is caused by the retinal activity following the onset of the light. (Arduini and his co-workers[6,7] have hypothesized and demonstrated that such an inhibition exists.) This inhibitory effect is evident in the decrease of deflection no. 1 (see Fig. XV-10). The facilitation of the cortical neurons following the flash is still evident in enhancement of the later deflections.

The mechanism of the enhancement effect is being investigated. Similarity of the waveforms for enhancement following the light onset and following a reticular stimulation suggests that the effects may involve the reticular formation. An important feature of the enhancement is that it disappears completely following deep anesthetization by Nembutal, which would not be at variance with this suggestion. However, there are findings that cast doubt upon a reticular involvement. First, there is the difficulty of obtaining the enhancement by reticular stimulation in the midpontine preparation, contrasted with the constancy and ease of the demonstration of enhancement following onset or termination of light. Second, the interval between light onset (or termination) and the time of delivery of the test shock for greatest enhancement is considerably shorter than the corresponding interval between the onset of the train of shocks to the reticular formation and the test shock. Finally, it appears that the enhancement following reticular stimulation is diminished more by a very light dosage of barbiturate anesthetic than the enhancement produced by transients in the light. None of these findings provides conclusive evidence one way or the other.

Further studies include a detailed investigation of the enhancement as a function of the time between the light onset (or termination) and delivery of the test shock. The question whether the facilitation of cortical neurons is limited to the primary visual area is also being investigated.

A. Cavaggioni, M. H. Goldstein, Jr., Eleanor K. Chance

References

1. G. H. Bishop and J. O'Leary, Potential records from the optic cortex of the cat, J. Neurophysiol. 1, 391-408 (1938).

2. F. Bremer and N. Stoupel, Analyse oscillographique comparée des réponses des aires de projection de l'écorce cérébral du chat, Arch. ital. Biol. 95, 1-19 (1957).

3. F. Bremer and N. Stoupel, Facilitation et inhibition des potentials évoqués corticaux dans l'éveil cérébral, Arch. internat. Physiol. Bioch. 67, 240-275 (1959).

4. S. Dumont and P. Dell, Facilitation réticulaire des mécanismes visuels corticaux, EEG Clin. Neurophysiol. 12, 769-796 (1960).

5. H. T. Chang, Cortical responses to stimulation of lateral geniculate body and the potentiation thereof by continuous illumination of the retina, J. Neurophysiol. 15, 5-26 (1952).

6. A. Arduini and T. Hirao, Enhancement of evoked responses in the visual system during reversible rètinal inactivation, Arch. ital. Biol. 98, 182-205 (1960).

7. A. Arduini and M. H. Goldstein, Jr., Enhancement of cortical responses to shocks delivered to lateral geniculate body. Localization and mechanism of the effects, Arch. ital. Biol. 99, 397-412 (1961).

E. STUDY OF THE HANDWRITING MOVEMENT

An experimental system to investigate the detailed properties of pen displacement and velocity as a function of time during handwriting has been put into operation. A commercial handwriting transmitter (Telautograph) has been modified so that its output signals, voltages from two precision potentiometers mechanically coupled to the pen, may be sampled, converted to digital form, and written on magnetic tape by the TX-0 computer. The analysis of these data is carried out by the IBM 7090 computer of the Computation Center, M.I.T.

Data analysis proceeds as follows. The x and y coordinate sequences are first recovered by solving the equations of mechanical constraint of the pen and then differentiated to yield the pen velocity vector function. A segmentation into strokes is carried out by using the zeros of the y velocity as segmentation points. A parametric description of the strokes thus obtained is determined on the basis of the following models.

1. Model A

This model matches half-wave sinusoidal segments $\dot{x}_m(t)$, $\dot{y}_m(t)$ to the experimentally obtained functions $\dot{x}(t)$ and $\dot{y}(t)$. It corresponds closely to that of Eden[1] with the

exception that his slope parameter σ has been replaced by a phase-shift parameter ϕ for mathematical convenience.

Approximating Equations

$$\left.\begin{aligned} \dot{x}_m(t) &= \dot{X} \sin[\omega(t-t_0)+\phi] + \dot{X}_c \\ \dot{y}_m(t) &= \dot{Y} \sin \omega(t-t_0) \end{aligned}\right\} \quad t_0 < t \leq t_2$$

Constraints

$$\omega = \pi/(t_2-t_0) \tag{1}$$

$$\dot{x}_m(t_0) = \dot{x}(t_0) \tag{2}$$

$$\dot{x}_m(t_2) = \dot{x}(t_2) \tag{3}$$

$$\int_{t_0}^{t_2} \dot{x}_m(t)\, dt = \int_{t_0}^{t_2} \dot{x}(t)\, dt \tag{4}$$

$$\int_{t_0}^{t_2} \dot{y}_m(t)\, dt = \int_{t_0}^{t_2} \dot{y}(t)\, dt, \tag{5}$$

where t_0 and t_2 are the experimentally determined segmentation points.

2. Model B

This model is an extension of Model A which allows for differences between the acceleration and deceleration sections of any stroke. The x-velocity amplitude, the frequency, and the phase shift between x and y velocities are allowed to take on different values over the two ranges.

Approximating Equations

$$\begin{aligned} \dot{x}_m(t) &= \dot{X}_1 \sin[\omega_1(t-t_0)+\phi_1] + \dot{X}_c && t_0 < t \leq t_1 \\ &= \dot{X}_2 \sin[\omega_2(t_2-t)-\phi_2] + \dot{X}_c && t_1 < t \leq t_2 \\ \dot{y}_m(t) &= \dot{Y} \sin[\omega_1(t-t_0)] && t_0 < t \leq t_1 \\ &= \dot{Y} \sin[\omega_2(t_2-t)] && t_1 < t \leq t_2 \end{aligned}$$

Constraints

Equations 2-5 of Model A and

$$\omega_1 = \pi/2(t_1-t_0)$$

$$\omega_2 = \pi/2(t_2-t_1)$$

$$\dot{X}_1 \cos \phi_1 = \dot{X}_2 \cos \phi_2 \qquad \text{(continuity at } t=t_1\text{)}$$

$$\omega_1 \dot{X}_1 \sin \phi_1 = \omega_2 \dot{X}_2 \sin \phi_2 \qquad \text{(continuous derivative at } t=t_1\text{)},$$

where t_0 and t_2 are experimentally determined segmentation points, as above, and t_1 is mathematically unconstrained but experimentally determined by minimizing the rms error between the experimental and matching functions.

3. Discussion

Writing samples produced by two input methods were obtained from one subject and processed. The first utilized the handwriting transmitter mentioned above in which the writing is executed on a horizontal surface (Input 1). In the second the writing is executed on the vertical face of the TX-0 computer oscilloscope (Input 2). The rms error ratios (rms error/rms signal) have been calculated to obtain an indication of the matching accuracy for both velocity and displacement functions and are given in Table XV-1. The extension of these results to include data from different subjects is under way at the present time.

Table XV-1. Matching accuracy of handwriting models.

Model		Input 1[a]	Input 2[b]
A	Displacement Error	.047	.046
A	Velocity Error	.232	.112
B	Displacement Error	.039	.045
B	Velocity Error	.175	.088

[a]Average of 9 samples

[b]Average of 4 samples

It is immediately apparent that a relatively poor match in the velocity domain will, when integrated, result in a quite acceptable match in the displacement domain. Hence, for Input 1, the input method of major interest, while Model A can be considered adequate in the displacement domain, it appears inadequate for the study of the detailed

intrastroke kinematics that are observable only in the velocity domain. Model B clearly effects a significant improvement for the velocity error over Model A. The differences between the two models for Input 2, for which the major part of the writing movement is executed by the elbow and the upper arm rather than the wrist and fingers, are much reduced and thus indicate that the kinematics of the segments of positive and negative acceleration within any stroke are much more alike. The smaller errors obtained for Input 2 indicate that this mode of writing is smoother and corresponds more closely to a harmonic oscillator system.

For Model A five parameters $[\dot{X}, \dot{X}_c, \dot{Y}, \omega, \phi]$ and for Model B eight parameters $[\dot{X}_1, \dot{X}_2, \dot{X}_c, \dot{Y}, \omega_1, \omega_2, \phi_1, \phi_2]$ are required to be known for the regeneration of any stroke. In each case the validity of the approximation is revealed by the close match between the original and regenerated two-dimensional handwriting patterns. The representations also achieve a reduction in the information storage per word over function samples at the minimum sampling frequency for perfect signal recovery if the velocity function is considered bandlimited at 20 cps.

The usefulness of the models for recognition work is primarily determined by the statistical properties of the above-given parameters. Experimental studies of the variations in stroke parameters with writer and context, as well as their consistency in the absence of such variations, are now being carried out. It is desired to determine the conditional probability densities $p(v_j/s_i)$, where v_j is the j^{th} parameter and s_i is the i^{th} stroke group, and strokes are grouped according to topological similarities. If the information content of the parameters is sufficiently high, that is, if $p(v_j/s_k) \gg p(v_j/s_1)$ for all $k \neq 1$, then reliable recognition of which stroke group the sample belongs to may be expected. The strong statistical constraints in sequences of particular strokes are to be used in the final identification of the individual letters.

P. Mermelstein

References

1. M. Eden, Handwriting and pattern recognition, IRE Trans., Vol. IT-8, pp. 160-166, 1962.

F. STROKE ANALYSIS OF DEVANAGARI CHARACTERS

Recent studies[1] of the characteristics of English cursive handwriting have raised the question whether writing systems unrelated to English might be a useful complementary study. For this purpose, the Devanagari script of India offers interesting possibilities because in its present form it has been fairly static for more than a thousand years, despite the development of closely related scripts, such as Bengali and Gujarati, which are much faster to write.

It is only in the past hundred years that mechanical methods of writing have become widespread in the West; and during the past decade strenuous efforts have been made to develop a satisfactory typewriter for Indian languages. Meanwhile, typesetting equipment that can produce aesthetically satisfying printed material is also being sought. At the present time, only hand-set type fonts making use of 400-800 pieces of type come close to looking authentic, and such useful devices as linotype have basic structural limitations that prevent them from accommodating such a variety of symbols.

Devanagari writing (used for Sanskrit, Hindi, Marathi, Punjabi, and in modified form for Gujarati and Bengali) is syllabic. That is, the unit of writing is the consonant (or consonant cluster) combined with the following vowel, which is written to the left, right, above or below. This combination is called an akshar. For example,

Vowels have a special written form when they appear without a preceding consonant.

1. Research Objectives

The aim of this research, then, is to study the structure of the Devanagari symbols with a view toward understanding whether and how the consonants and consonant clusters (a brief survey uncovered 221 distinct clusters) can be generated, and described, distinguished or selected, in a consistent manner.

A consonant (or cluster) appearing alone is read as containing the "intrinsic" vowel, /ə/. Frequency studies in Hindi and Marathi show that the most frequent occurrences are of a single consonant with /ə/, and therefore we shall study the writing of the 33 Devanagari consonants first.

In this report a stroke analysis appropriate to Devanagari is presented, together with the joining rules for normal writing. These rules form a complete set with respect to generation of the consonants, as shown in Fig. XV-11.

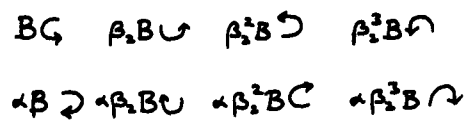
2. Development of the Set of Basic Strokes

In contrast to English cursive script, penlifts occur in Devanagari even in the writing of a single "letter." Strokes are written from left to right and from top to bottom. There are very few exceptions. It is possible to develop most of the strokes by applying two kinds of transformations to a basic Straight Stroke ↓ /A/ and Curved Stroke ↄ /B/. We call the operator Reflection (about a vertical), α. We call the operator Rotation, β (Rotation by 45°, β_1; Rotation by 90°, β_2). Applying β_1 and β_2 to /A/, we get

$$A \downarrow \beta_1 A \searrow \beta_2 A \rightarrow \beta_1\beta_2 A \nearrow \beta_2^2 A \uparrow \beta_1\beta_2^2 A \nwarrow \beta_2^3 A \leftarrow \beta_1\beta_2^3 A \swarrow$$

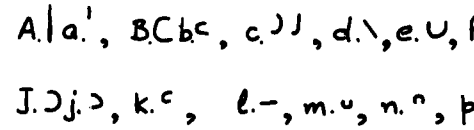
Of these, only the first four occur. They are labelled (for convenience, and to avoid

subscripts): /A/, /D/, /L/, /Q/. Applying α and β_2 to /B/, we get



Of these, all occur except the last. They are labelled (again, for convenience): /B/, /E/, /J/, /N/, /H/, /M/, /K/. Two other strokes  /f/ and  /g/ are reflections of each other. Two miscellaneous strokes  /C/ and  /P/ complete the set of 15 Basic Strokes.

The stroke sizes have been quantized to two levels. For the larger, we use upper-case labels; and for the smaller, we use lower-case labels. (Concerning the third Basic Stroke see II (iii) below.)



For three strokes, both large and small sizes occur. In all other cases only the smaller size is utilized.

A distinctive feature of Devanagari is the top line that "ties" together the character (and is written last). Certain characters have a broken top line, and for these /(r)/ is added to the linear-form description of the character.

One other symbol, /s/, is used to signify a penlift. Where this is redundant information (the penlift being due to the direction in which the previous stroke was written), /$\not$/ is used.

3. General Comments on Characteristic Features of the Strokes

We shall list some characteristic features of the strokes in a purely descriptive manner. Then we shall discuss some cases in which contextual clues are relevant to the role of the individual stroke in forming the characters. Relatively simple rules suffice to indicate whether the larger or smaller size of a given stroke is to be used in a given situation. Thus the same code could unambiguously apply to both. Similarly, in certain contexts a variant on the usual form (or shape) of a stroke is used. To simplify reading, we have retained the lower- and upper-case designations in the "linear forms" given in Fig. XV-11.

I. Characteristic Features of the Strokes

(i) Written from top to bottom: A B c d F g h
Written from left to right: d e l p q
Written from right to left: c F m n
Written from bottom to top: J k q

(ii) Initial movement downward: A c d e m p
Initial movement to right: d g h J p q
Initial movement to left: B F k
Initial movement upward: n q

(iii) Requiring counterclockwise movement: B e f j m p
Requiring clockwise movement: c g h n k

(iv) Positioning in letter space: Height
Always found at same distance below top line: A a B c F G J k l m p
Height in character can vary: b d e h n q

The rarity of strokes written leftward or upward, as indicated in I (ii), is even more striking when the relative frequency of these strokes is taken into consideration.

The data of I (iv) indicate a high degree of redundancy, in that most strokes are not found except in certain regions of the letter space and may perhaps explain the readability of the script despite its complexity.

II. Notes on Certain Strokes

(i) /A,a/ Both large and small forms are always written from the top line
As an initial stroke, it occurs only as /a/
As medial or final stroke, it occurs only as /A/

It is interesting to note that none of the characters has a vertical line segment except that hanging from the top line.

(ii) /B,b/ As an initial stroke this occurs only as /b/, starts from the top line only in sequence /bb../. Otherwise it is written in the middle of the letter space.
In medial position the form is /b/ in context /abh/; otherwise B as in aB, aBJ, aBn, aBjc.

(iii) /c/ This stroke has two forms. It is written ( ) when followed by /l/; otherwise it is written ( ).

(iv) /e/ In context /le../ the form is ( ) instead of the normal ( ).

(v) /J,j/ In context /aBJ/ the form is ( ); i.e., /J/. Otherwise ( ), /j/.

(vi) /l/ Always written at mid-height in letter space. The group /..lA../ is frequent and is uniformly written .

4. General Rules of Formation of Characters from Strings of Strokes

The rules presented below may be considered a preliminary set. As formulated at present, they deal with the construction of the consonants. They have been left flexible, however, in order to accommodate the numerous and complex cases of conjunct characters. A significant feature of the rules is that transitional strokes are obviated except with /A/.

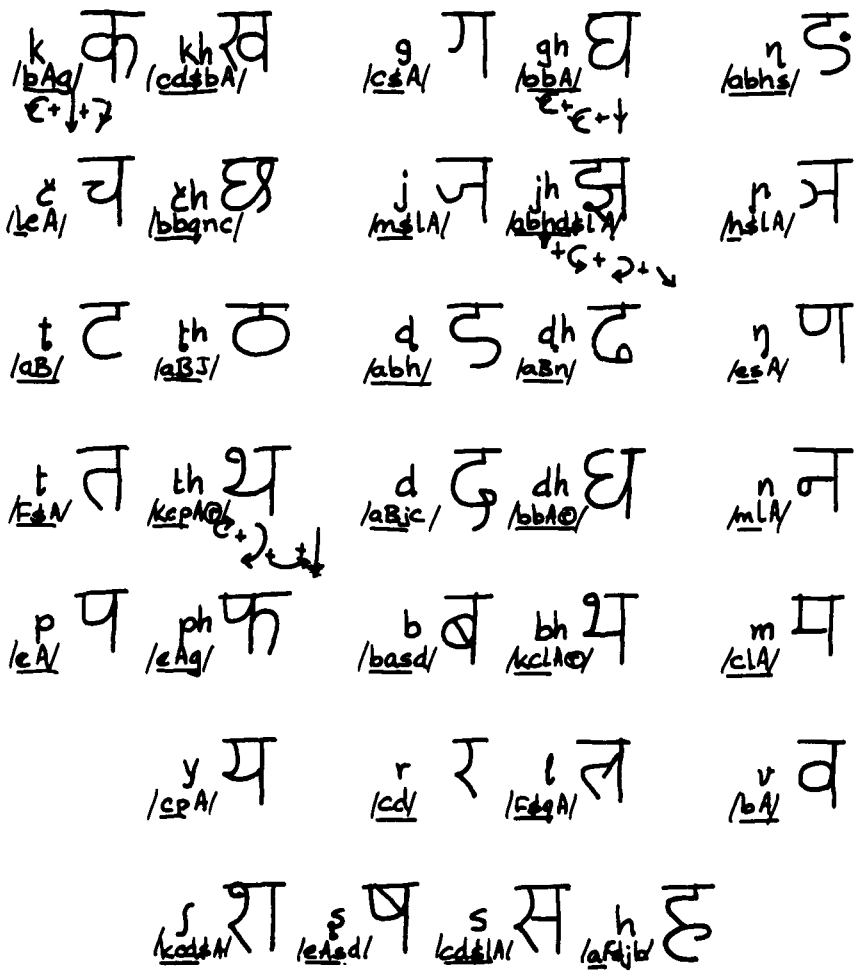


Fig. XV-11. The 33 Devanagari consonants and their "linear" representation.

III. General Rules

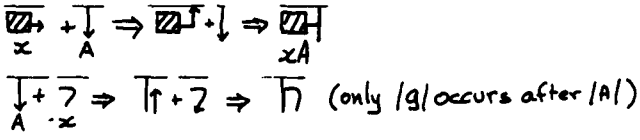
(i) The highest point of the first stroke is always at the top line, except that initial /b/, /F/, /h/, /l/, /m/ are written at mid-height. (Note that initial /bb../ is written from the top line.)

(ii) Strokes always commence from the end point of the previous stroke except as in III (v) below.

(iii) Mark /#/ implies a penlift that is due to the direction of stroke, and is a redundant instruction. Mark /s/ implies a penlift that is due to the fact that the character has two segments, and implies a space left between the segments.

(iv) After the listed strokes comes the conventional top line. In cases in which the additional instruction /(r)/ is added, it implies that the top line is written over the right half of the character only.

(v) Added transitional strokes are used before and after /A/ as follows:



IV. Examples of Operation of the Rules

/b: * /bb: ε /bA/: ɗ /bbq: ɛ/ /bbqn: ɛ9 /bbqnc: ɛ8 /bbqnc/: ɛ8

In the first example, * implies uncertainty till the sequence is known.

5. Remarks

The work reported here provides a means of uniquely generating the 33 consonants of the Devanagari alphabet from a set of 15 Basic Strokes. In Fig. XV-11 the stroke descriptions, which are essentially linear representations of two-dimensional characters, are underlined to the point where unambiguous identification is possible.

This work can be logically extended to the study of akshar formation and the adaptation of the present strokes and rules to this far more complex situation. In another direction lies the comparison of the script described in this report, as used in Sanskrit, Hindi, Marathi, and Punjabi, with related scripts. Gujarati on the West coast of India, and Bengali in the East, have developed written alphabets that are obviously offshoots of Sanskrit-Devanagari. The study of these scripts in relation to their parent writing form is not only of anthropological interest but also provides an opportunity for a comparative study of the cognitive roles of the constituent strokes in the related-but-different formations.

J. G. Krishnayya

References

1. M. Eden and M. Halle, The Characterization of Cursive Handwriting, Information Theory—Fourth London Symposium, edited by C. Cherry (Butterworths Scientific Publications, Washington, D. C., 1961).

XVI. NEUROPHYSIOLOGY*

W. S. McCulloch
M. A. Arbib
F. S. Axelrod
M. Blum
J. E. Brown
R. C. Gesteland
M. C. Goodall
W. L. Kilmer
K. Kornacker
J. Y. Lettvin
Diane Major
L. M. Mendell
N. M. Onesto
W. H. Pitts
J. A. Rojas
A. Taub
P. D. Wall

A. THE EFFECT OF PENTOBARBITAL ON THE VISUAL SYSTEM OF RATS

The optic nerves of albino rats anesthetized with sodium pentobarbital (50 mg/kg) were exposed by removing the cerebrum lying over the nerves by suction. In such a preparation, the circulation in the retina and optic nerve remains intact. The dura and nerve sheath were removed from the nerve between the optic foramen and the optic chiasm and single units were recorded by means of metal-filled glass micropipettes, plated with platinum black.[1] Single units can be stably recorded for one hour and longer; moreover, units recorded in sequence along the track of the electrode can be recovered upon withdrawal of the electrode in the reverse sequence.

In such a preparation under pentobarbital the most striking activity recorded is a synchronous, spontaneous bursting of most of the active units at approximately 15 cps (Figs. XVI-1 and XVI-2). Almost all such units respond with a burst of spikes to the "off" of the general illumination, and are desynchronized when the general illumination is turned on (Fig. XVI-2). This bursting "off" unit activity is quite well synchronized through the nerve; a 5-micron KCl pipette, placed on the surface of the nerve, records a periodic potential of approximately 15-cps frequency (Fig. XVI-3).

The bursting "off" unit activity is not present in the optic nerve axons of an animal prepared without any barbiturate; these animals were prepared under ether anesthesia. A midcollicular section was performed, followed by removal by suction of the cerebrum overlying the optic nerve. If the optic nerve in such an unanesthetized animal is crushed close to the optic chiasm and records taken distal to the crush, no such bursting activity is seen. This suggests that the bursting activity is not due to suppression of some efferent activity by the barbiturate. Furthermore, an electrode records many more units in the unanesthetized animal, many of which are "on" units. If the sequence and number of units along an electrode track is noted and Nembutal administered, as the level of anesthesia deepens and the electrode is withdrawn, the nerve is found to be relatively much more quiet, with groups of bursting "off" units not before observed now being recorded, which have quiet regions between them. Conversely, as the level of anesthesia lightens,

*This work was supported in part by Bell Telephone Laboratories, Inc.; The Teagle Foundation, Inc.; the National Institutes of Health (Grant NB-01865-05 and Grant MH-04737-02); and in part by the U. S. Air Force (Aeronautical Systems Division) under Contract AF33(616)-7783.

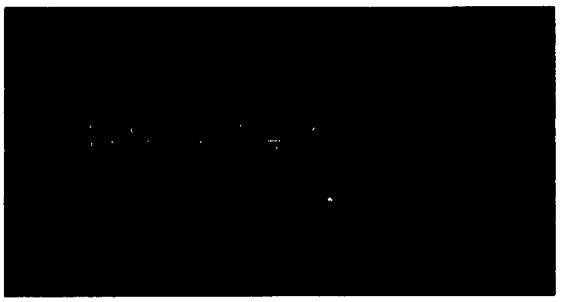


Fig. XVI-1. A single unit of the "bursting off unit" variety recorded in the optic nerve of an albino rat under pentobarbital. Time marker: 50 per second.

Fig. XVI-2. Several bursting units recorded in the optic nerve of an albino rat under pentobarbital. The background illumination was turned on in the middle of the sweep and off at the end. Time marker: 2 per second.

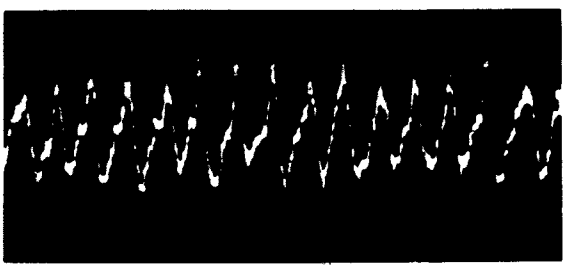


Fig. XVI-3. Slow potentials recorded with a gross electrode from the surface of the optic nerve of an albino rat under pentobarbital. Time marker: 4 per second.

the bursting activity diminishes and types of units other than the bursting "off" units are recorded.

This bursting "off" unit activity was also seen in albino rats under allobarbitone (Dial, Ciba), and in both albino and Long-Evans hooded rats under pentobarbital with the recordings taken stereotaxically with the same type of low-impedance electrodes.

J. E. Brown, J. A. Rojas

References

1. R. C. Gesteland, B. Howland, J. Y. Lettvin, and W. H. Pitts, Comments on microelectrodes, Proc. IRE 47, 1856 (1959).

B. SINGLE-UNIT RESPONSES IN THE CEREBELLUM OF THE FROG

The cerebellum of the frog is a bean-shaped organ lying on the dorsal side of the brain between the medulla and the tectal lobes. The largest nerve-fiber bundles entering the cerebellum are the (dorsal) spinal tract and the vestibular tract. Small tracts course in from nerves IV and V, from the tectum, and from sundry nuclei in the medulla. The histological structure of the organ upon which these fiber tracts of different sensory modalities converge is similar to, but less complex than, that of the mammal. That is, according to the scanty and sketchy anatomy that we find (in the frog) there are the three distinct layers – the molecular, the granular, and the Purkinje cell layers – but their arrangement and interconnection is less elaborate than in the mammal.[1,2]

The present study attempts to reveal through microelectrode recording how each modality is handled separately and in combination by short-term responses in cells of the cerebelium. The recording was done with metal-filled, platinum-plated, glass electrodes on curarized (in several cases, uncurarized) frogs (R. pipiens). I distinguished the stimuli applied to the frog mainly by general qualitative differences.

In harmony with the nature of the stimuli, this report is a qualitative description of the behavior of cerebellar cells in the frog. This description is not exhaustive and deals only with those cells in which the response was a clear-cut change in activity. It will be presented as a set of categories representing the nature of the applied stimuli – vestibular, mechanical, visual, and combinations of these three.

Some of the vestibular organs are simultaneously stimulated by a rotation of the frog about an axis inclined to the horizontal. Cells that respond to this mode of stimulation are of the following types:

(i) One kind has a periodic resting rate, responds to rotation both clockwise and anticlockwise with a continuous barrage, the frequency of which depends on the velocity of rotation. These resting rates change with changes in the position of the head.

(ii) Another kind usually has an aperiodic, bursty, resting rate and responds to

rotation in one direction with a burst whose frequency depends on the magnitude of the acceleration. It is inhibited by, or responds only weakly to, a rotation in the other direction. Such a cell shows a rapid adaptation to successive rotatory (back and forth) stimuli. It usually does not show nicely periodic changes in resting rate with changes in position of the head; although, in several cases, a change in position did produce a quickening of the rate (Fig. XVI-4).

(iii) Another kind is (a) silent and responds to a rotation in one direction with an abrupt burst of five or more spikes whose frequency is proportional to the rate of change of acceleration, or (b) is silent and responds with a spike or two at the end of rotation. Both (a) and (b) accommodate to a second stimulus presented with a period of 10 seconds.

In all of the above-mentioned cells possessing a resting rate, that resting rate is depressed after a response but slowly returns after a period of at least 10 seconds (often 1 minute). The length of this period depends on the strength and the duration of the stimulus.

Mechanical stimulation is performed on the limbs, and to a lesser extent on other parts of the body – neck, back, snout, eye muscles, cornea. Cells that respond to this mode of stimulation are of the following types:

(i) One has usually an aperiodic resting rate, responds with a burst to changes in limb position, and rapidly and completely adapts to a second stimulus for a period of at least 10 seconds. These cells are sensitive to at least one and perhaps several muscles acting about a particular joint in the limb, and often will respond to the same stimulus in both left and right limbs (Fig. XVI-5).

(ii) Another has a periodic resting rate and responds with a burst to any one of the following stimuli, the choice depending on the cell: Pressing on the limb pads (of sometimes one, two, or all four limbs), pressing on the eye muscles, and blowing and brushing on the cornea and various parts of the skin. These stimuli depress the periodic resting rate after a response.

(iii) A third is silent and gives a periodic response whose frequency is proportional to the amount of bend of a particular joint in one limb. These responses are nonadaptive and may reach a high firing rate of several hundred spikes per second, and maintain this rate for as long as the joint is bent (Fig. XVI-6).

The only other sensory modality seen thus far to affect some cells of the frog cerebellum is vision. These cells have an aperiodic resting rate and respond with an abrupt burst to the movement of an object (for example, a hand) within the visual field of an eye. They adapt after a single stimulus, and will not respond again to this stimulus for at least 15 seconds. They thus have the same properties as the 'newness' neurons seen by Lettvin and others in the optic tectum.[3]

Interactions between the various modalities occur in the cells described above.

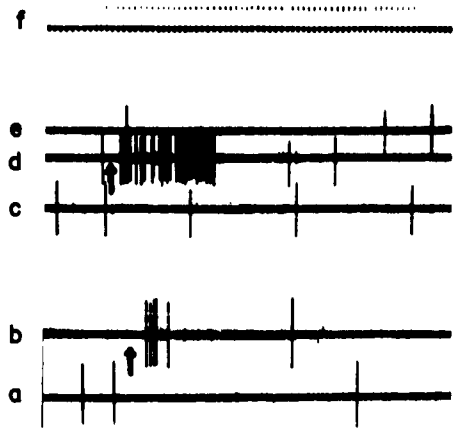


Fig. XVI-4. Vestibular type ii cell. Onset of stimulus indicated by an arrow beneath the trace. Spike height ~1.5 mv.

a. resting rate (head in normal position).
b. rotation 90° counterclockwise.
c. resting rate in new position.
d. rotation 90° clockwise.
e. initial resting rate.
f. time base, 10/sec.

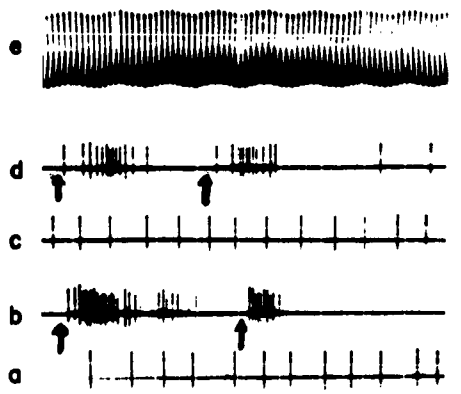


Fig. XVI-5. Mechanical type i cell. Spike height ~2 mv.

a. resting rate.
b. right leg picked up and put down (up and down indicated by arrows).
c. resting rate.
d. left leg picked up and put down.
e. time base, 10/sec.

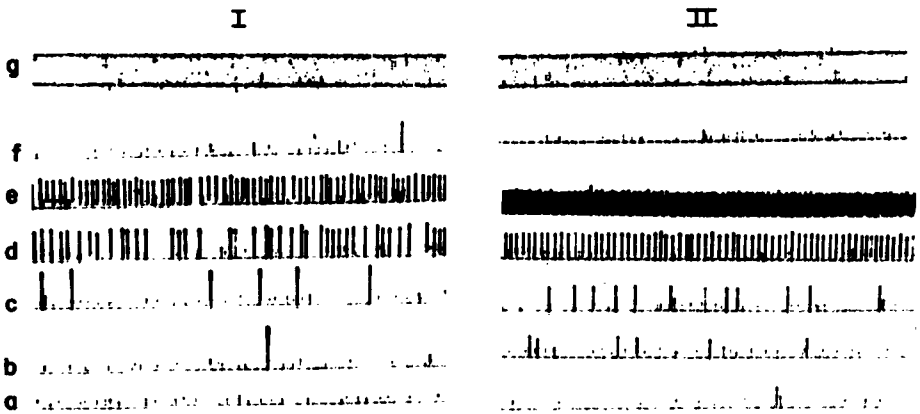


Fig. XVI-6. Mechanical type iii cell. Spike height ~1 mv.

Part I. Extension about ankle joint with knee joint in the normal position.
a. resting rate.
b.-d. increasing extension about joint.
e. maximal extension.
f. resting rate.
g. time base, 10/sec.

Part II. Extension about ankle joint with a fixed flexion about knee joint.
a. resting rate.
b. increasing extension about ankle joint (increase in the same steps as in Part I. b-d).
e. maximal extension.
f. resting rate.
g. time base, 10/sec.

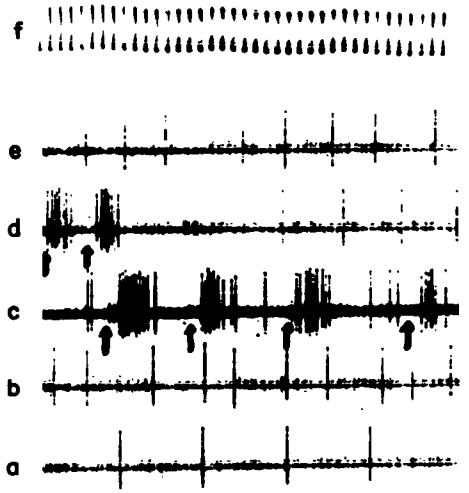


Fig. XVI-7. Interaction type a cell. Spike height ~1 mv.

a. resting rate (head in normal position).
b. resting rate (head rotated 90° clockwise).
c. left toe pad pressed on four successive times.
d. left toe pad pressed on twice (note depression of the resting rate).
e. resting rate (head position as in b).
f. time base, 10/sec.

These cells behave as mixtures of the following kinds:

(a) Vestibular 1 and Mechanical 2. These cells have a periodic resting rate and will respond to rotation and change their resting rate with changes in the position of the head, will respond, for example, with a burst to pressing on left or right toe pad, and will have their resting rate depressed immediately thereafter (Fig. XVI-7). Similar responses are elicited in other cells with other mechanical stimuli.

(b) Vestibular 2 and Mechanical 1. These cells have an aperiodic resting rate and will respond with a burst to either rotation or movement of one or several limbs, and will have their resting rate depressed immediately thereafter.

(c) Vestibular 2 and Visual. These behave as those in (b), except that vision rather than mechanical stimulation causes response. Visual responses have been seen in only this type of cell.

No attempt has yet been made at strict anatomical localization of the above-mentioned groups of cells with respect to the layers of the cerebellum. Physiologically, in penetrating the dorsal surface of the cerebellum, the electrode, inclined caudally at a very slight angle to the vertical, encounters three distinct layers consisting of the following groups of cells:

(i) vestibular 2, vestibular 3, mechanical 1, visual

(ii) vestibular 1, mechanical 2

(iii) mechanical 3.

The time constants of the cells in these three layers are also appreciably different. The height of the potentials recorded in the first layer is invariably greater than that of those in the other two, and the cells in this layer are electrically closely packed. These cells are probably the Purkinje cells, which are the largest (~10 μ in diameter) and the most closely packed cells in the cerebellum.

With the exception of vestibular 3 responses, the short-term response pattern of the cerebellar cells is not appreciably different from that found in the fiber tracts entering the cerebellum. In fact, the response in mechanical 3 seems no different from what would be expected in a tendon-stretch receptor itself, unless, indeed, this response is combinatorial with other stretch receptors in the limb (as is perhaps the case in Fig. XVI-6). The differences are found in the combination of the responses of various modalities and in that of the responses from several limbs within a single cell.

The most patent use to the frog of the short-term responses of cerebellar cells is in the rapid control of efferents to muscle. For example, Granit, Holmgren, and Merton[4] have shown that in the cat the cerebellum in some way regulates the proportion of the two routes of muscle excitation, α and γ, in use at any time. That there is a large representation of the γ efferents in the cerebellum is seen in the large decrease in cellular activity in layer ii when a frog is curarized. (Curare knocks out muscle spindle activity.) The proportion of the two routes in use in a given muscle at any time

is very likely dependent on vestibular, dermal, visual, and other muscular activity. The nature of these short-term responses allows the cerebellum to perform just such a regulation.

The long-term responses of the cerebellar cells consist in the setting and modulation of the complex resting rates (not strictly aperiodic but almost cyclic in the repetition of a bursting pattern) of those cells by previous vestibular, mechanical, and perhaps visual stimuli. These kinds of responses are suited to making the cerebellum in part an organ that can 'remember' previous trajectories of a frog in its milieu, and that can plot intended trajectories of the animal. This bit of speculation will be the object of further study on the cerebellum of the frog.

F. S. Axelrod

References

1. O. Larsell, The cerebellum of the frog, J. Comp. Neurol. 36, 89-113 (1923).

2. P. Glees, C. Pearson, and A. G. Smith, Synapses on Purkinje cells of the frog, Quart. J. Exptl. Physiol. 43, 52-61 (1958).

3. J. Y. Lettvin, H. R. Maturana, W. H. Pitts, and W. S. McCulloch, Two remarks on the visual system of the frog, in Sensory Communication, edited by W. A. Rosenblith (The M. I. T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York and London, 1961).

4. R. Granit, B. Holmgren, and P. A. Merton, The two routes for excitation of muscle and their subservience to the cerebellum, J. Physiol. 130, 213-225 (1955).

XVII. NEUROLOGY*

L. Stark
F. H. Baker
R. W. Cornew
H. T. Hermann
J. C. Houk, Jr.
F. Naves
T. Rowe
A. A. Sandberg
Susanne Shuman
J. I. Simpson
Gabriella W. Smith
I. Sobel
S. F. Stanten
A. Troelstra
E. C. Van Horn, Jr.
G. L. Wickelgren
P. A. Willis
S. Yasui
L. R. Young
B. L. Zuber

A. BLACK-BOX DESCRIPTION AND PHYSICAL ELEMENT IDENTIFICATION IN THE PUPIL SYSTEM

Black-box input-output analysis of nonlinear systems is, at present, a control-systems

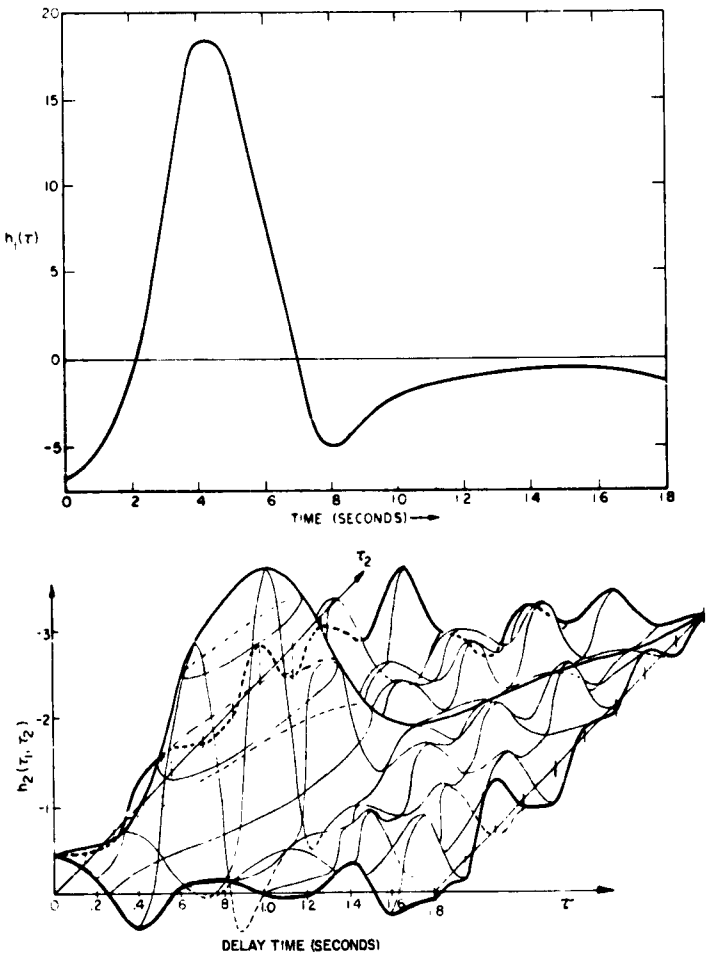


Fig. XVII-1.

First- and second-order kernels of open-loop pupil light servomechanism.

*This research is supported in part by the U. S. Public Health Service (B-3055-3, B-3090-3, 38101-22), the Office of Naval Research (Nonr-1841 (70)), the Air Force (AF33(616)-7588, AFAFOSR-155-63), and the Army Chemical Corps (DA-18-108-405-Cml-942); and in part by the National Institutes of Health (Grant MH-04734-02).

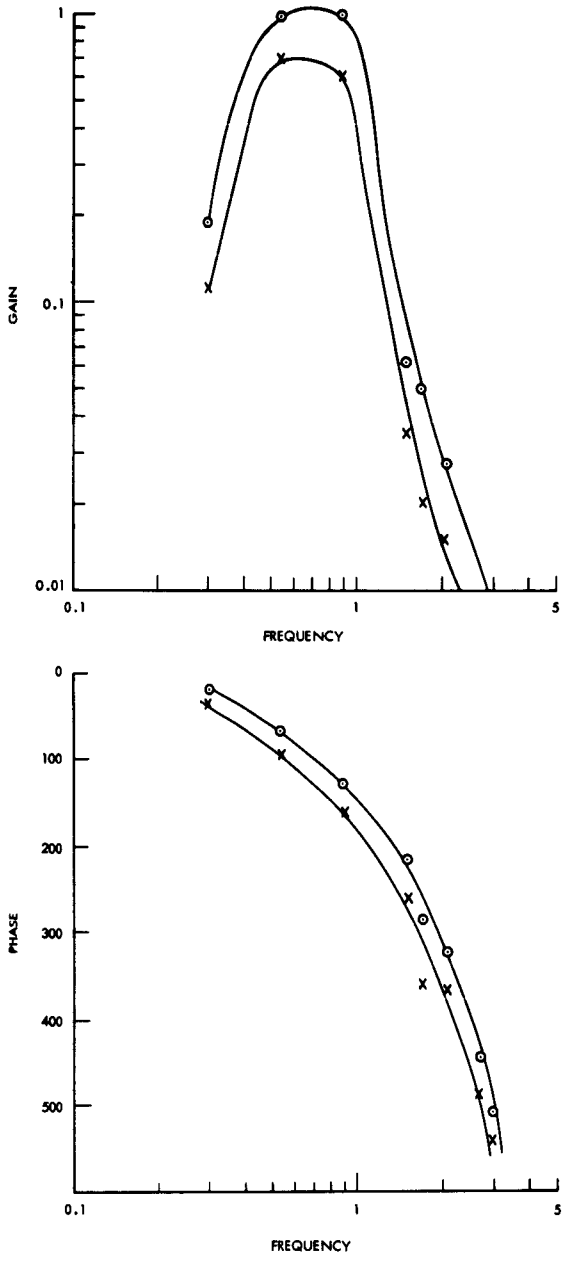


Fig. XVII-2. Bode plot of drugged (X) and undrugged (⊙) pupil.

area of much activity. Functional analysis[1-4] has a mathematical elegance that is attractive, and we have applied this method[5] to the pupil light reflex. Figure XVII-1 shows $h_1(T)$, the first-order kernel, and $h_2(T_1, T_2)$, the second-order kernel, as two-dimensional and three-dimensional functions, respectively. Work is now under way to refine the experimental method to obtain more consistent and reliable data.

However, since our goal is identification of the physical laws of the anatomical-physiological elements that together make up the pupillary system, we have attempted to dissect into the black box. The high-frequency cutoff is probably due to the output mechanical elements, the iris muscles, as the experiment illustrated in Fig. XVII-2 demonstrates.[6] Here, locally applied drugs that partly overstimulate both the sphincter and dilator muscles reduce the bandpass of the system and the gain. This confirms a previous experiment[7] in which gain measurements were not obtained.

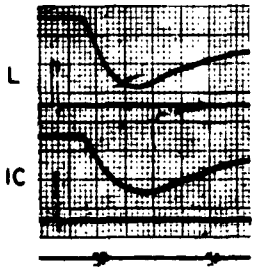


Fig. XVII-3. Impulse response of pupil to light (L) and electrical intracranial (IC) stimulation.

Frank H. Baker[8] has shown that in anesthetized cats direct intracranial stimulation by electrodes stereotactically placed near the motor fibers of the oculomotor nerve produces pupillary contraction with approximately the same transport delay and third-order response as does light stimulation of retina, as shown in Fig. XVII-3. Estimates of nerve conduction times and synaptic delays for the pupillary system range from 10 msec to 20 msec, in agreement with the assigning of 90 per cent of the transport delay and all of the high-frequency attenuation to the output neuromuscular elements.[9]

These experiments indicate the range of approaches utilizing system theory, and neurophysiological and neuropharmacological dissection techniques, both of which are necessary to make a quantitative and into-the-black-box analysis of a neurological control system.

L. Stark

References

1. N. Wiener, Nonlinear Problems in Random Theory (The Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1958).

2. G. Zames, Nonlinear Operators for System Analysis, Sc.D. Thesis, Department of Electrical Engineering, M.I.T., August 1960; Technical Report 370, Research Laboratory of Electronics, M.I.T., August 25, 1960.

3. Y. W. Lee and M. Schetzen, Measure of the kernels of a nonlinear system by crosscorrelation, Quarterly Progress Report No. 60, Research Laboratory of Electronics, M.I.T., January 15, 1961, pp. 118-130.

4. J. Katzenelson and L. A. Gould, The design of nonlinear filters and control systems, Information and Control 5, 108-143 (1962).

5. J. Katzenelson and A. A. Sandberg (unpublished experiments, 1961-1963).

6. J. W. Stark, The Effects of Drugs on the Transfer Function of the Human Pupil System, S.B. Thesis, Department of Biology, M.I.T., May 1962.

7. L. Stark and F. Baker, Stability and oscillation in neurological servomechanism, J. Neurophysiol 22, 156-164 (1959).

8. F. H. Baker (unpublished experiments, 1959-1963).

9. L. Stark, Environmental clamping of biological systems. The pupil servomechanism, J. Opt. Soc. Am. 52, 925-930 (1962).

B. PUPIL VARIATION AND DISJUNCTIVE EYE MOVEMENTS AS A RESULT OF PHOTIC AND ACCOMMODATIVE STIMULATION

There are several methods of stimulating the iris muscles which result in a variation of the pupil diameter (Fig. XVII-4). First, the eye can be stimulated with light. A signal goes from the retina to the central nervous system (CNS), and then back to the iris muscles.[1] This photic response is involuntary, and thus prediction does not occur. Second, a target can be moved on the optical axis of one eye, while a cooperative subject tries to keep it in focus. It is evident that this method of providing a stimulus necessitates a voluntary contrubution by the subject and thus prediction may occur. In this situation, an error signal goes to the CNS, indicating how far the image is out of focus. This is an even-error signal, but in our present experimental arrangement, since there are additional clues such as the size and brightness of the target, we actually work with an odd-error signal.[2] From the CNS there is a signal path to the ciliary muscle and the dioptric strength of the lens varies in such a way as to obtain clear vision. Also, an additional signal goes from the CNS to iris muscles, and pupil diameter varies to control depth of focus of the eye.

Third, the accommodation input by way of the CNS results in a disjunctive eye movement of the other eye, although this eye cannot see the target. Figure XVII-4 is a simplified block diagram of these various inputs, outputs, and interactions. The experimental arrangement is shown in Fig. XVII-5; it is possible to stimulate one eye with either a light or an accommodation input and to measure the pupil size. The associated disjunctive movements of the other eye which resulted from the

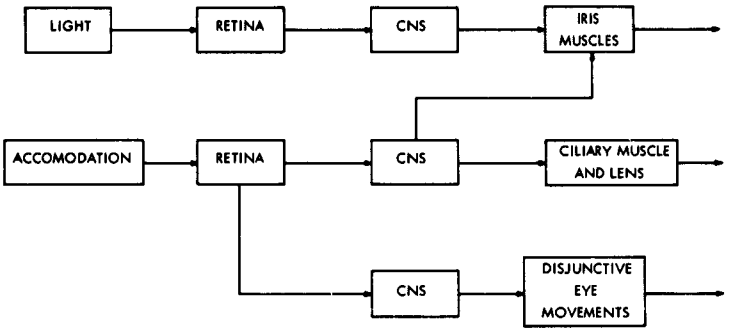


Fig. XVII-4. Block diagram of the interacting iris-lens-convergence system.

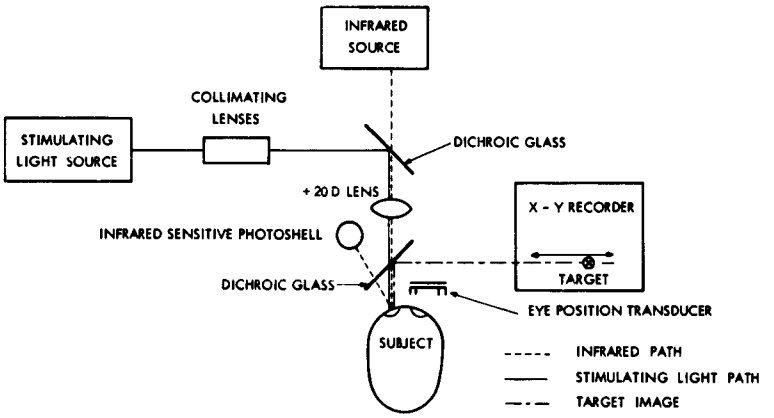
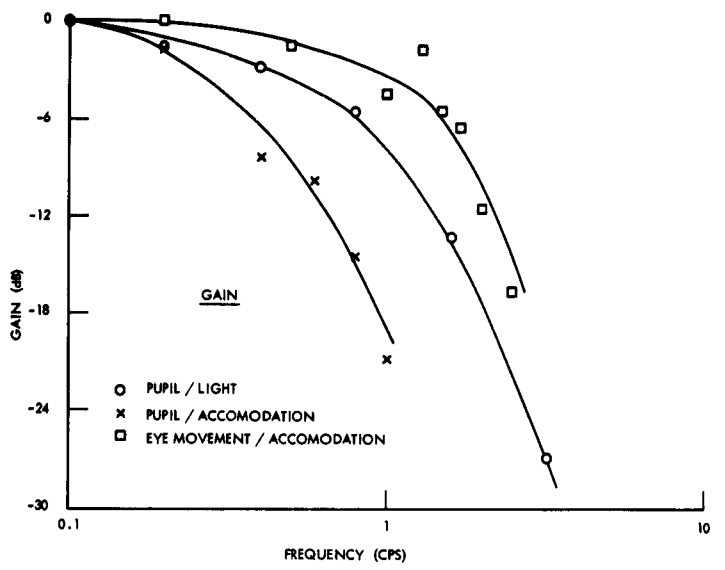
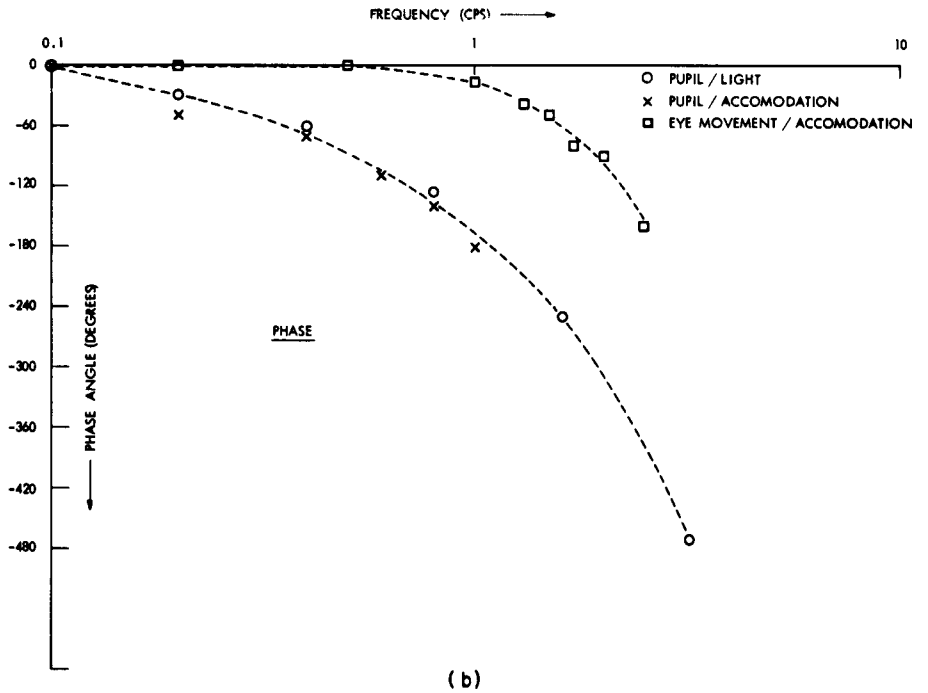


Fig. XVII-5. Experimental arrangement.



(a)



(b)

Fig. XVII-6. Frequency-response plots of various input-output relationships.

accommodation input were measured separately with equipment previously described.[3]

To obtain an idea of the dynamics of the various systems, the eye was stimulated with a sinusoidal light input or with a sinusoidal target-position input in the optical axis of one eye. The variations in pupil diameter of the same or the disjunctive movements of the other eye are measured. Bode plots are shown in Fig. XVII-6.

A. Troelstra, B. L. Zuber, J. I. Simpson, L. Stark

References

1. L. Stark, Y. Takahashi, and G. D. Zames, Dynamics of the human lens system, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 337-351.

2. L. Stark and Y. Takahashi, Accommodation tracking, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 205-212.

3. L. Stark and A. Sandberg, A simple instrument for measuring eye movement, Quarterly Progress Report No. 62, Research Laboratory of Electronics, M.I.T., July 15, 1961, pp. 268-270.

C. REMOTE PATIENT-TESTING INSTALLATION

We have set up a communication link between the Howe Laboratories of the

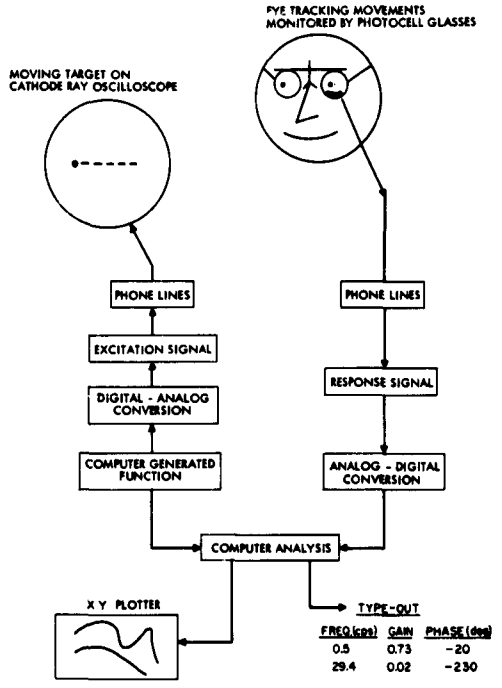


Fig. XVII-7. Block diagram of remote patient-testing installation using on-line computer.

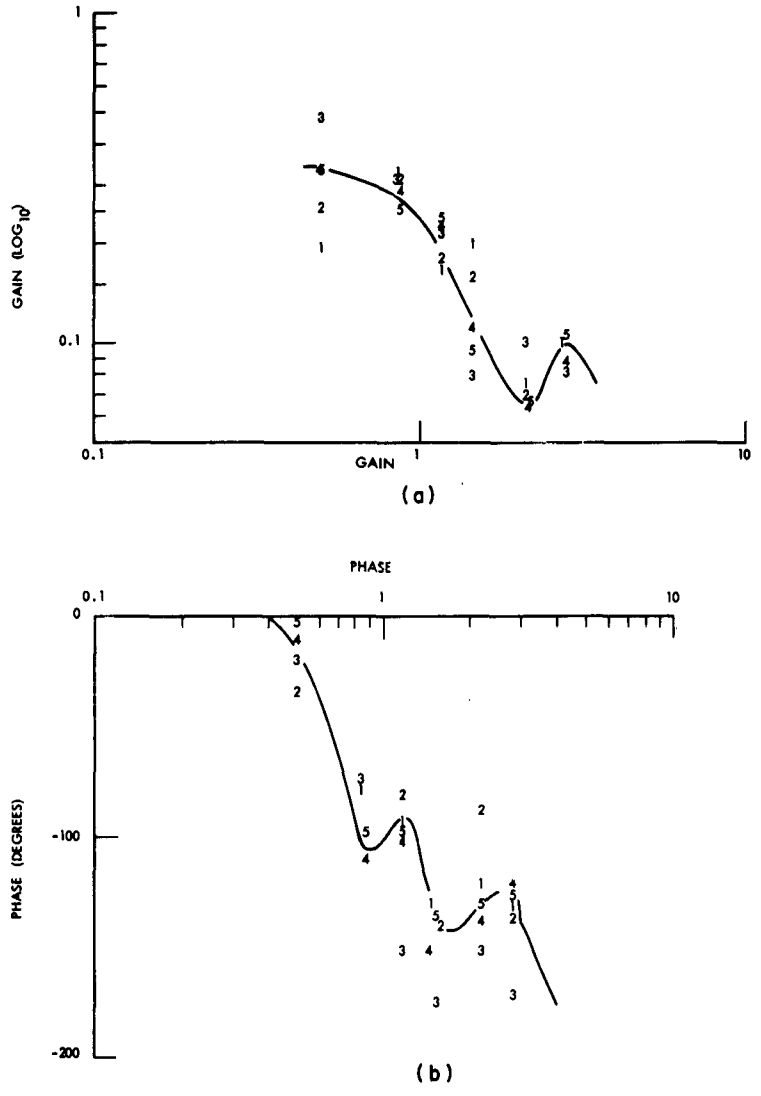


Fig. XVII-8. Bode plot of eye-tracking movement control system.

Table XVII-1. Gain and phase as a function of frequency typed out by on-line computer.

ANNE TROELSTRA TRACKING 1 18 62 PAW LS GS

Run No.

AVERAGE 000110

	FREQUENCY	GAIN	PHASE
	00.496	00.197	- 34.507
	00.859	00.328	- 74.035
1	01.160	00.169	- 91.111
	01.455	00.203	-124.507
	02.162	00.078	-122.619
	02.757	00.103	-137.103

END

AVERAGE 000119

	FREQUENCY	GAIN	PHASE
	00.496	00.261	- 36.869
	00.859	00.312	-111.339
2	01.160	00.183	- 83.970
	01.455	00.160	-138.365
	02.162	00.072	- 88.613
	02.757	00.101	-139.574

ANE TROELSTRA 1 18 62 EOM AT COGAN LABAVERAGE 000130

	FREQUENCY	GAIN	TPHASE
	00.496	00.486	03.580
	00.859	00.310	-128.197
3	01.160	00.218	-129.181
	01.455	00.080	-156.037
	02.162	00.101	-128.660
	02.757	00.083	-154.824

END

AVERAGE 000109

	FREQUENCY	GAIN	TPHASE
	00.496	00.337	- 10.619
	00.859	00.287	-115.173
4	01.160	00.222	- 99.144
	01.455	00.115	-150.253
	02.162	00.064	-138.814
	02.757	00.089	-123.566

ENDAVERAGE 000105

	FREQUENCY	GAIN	TPHASE
	00.496	00.324	- 03.578
	00.859	00.255	- 99.093
5	01.160	00.238	- 95.431
	01.455	00.095	-131.185
	02.162	00.068	-131.632
	02.757	00.101	-131.888

END

Massachusetts Eye and Ear Infirmary[1] and an on-line computer (GE 225) in the Electronic Systems Laboratory, M.I.T., for studying neurological eye-movement defects. The communication links are four pairs of standard low-fidelity "direct" telephone lines, traveling approximately 3 miles in roundabout fashion from Cambridge to Boston.

The experimental arrangement, as shown in Fig. XVII-7, consists of a pseudo-random excitation signal that is generated by the computer, converted by the computer DIOB (data input-output buffer) to an analog voltage, and transmitted over one pair of lines (0-2000 cps bandwidth) to drive a horizontally moving spot on an oscilloscope face. The patient is instructed to follow the spot; the angular direction of his eyes is measured by a pair of photocell goggles,[2] and this response signal is sent back over another pair of lines, and is digitalized by the DIOB. The GE 225 computer then analyzes the gain and phase lag at each frequency of interest and types out the information. A typical typeout is shown in Table XVII-1. Figure XVII-8 is a plot of five successive runs on A. Troelstra.

The peak in the frequency response has been predicted by a sampled-data model of the eye-movement control system.[3] An X-Y plotter will be installed in the hospital laboratory and the plotted frequency response will be available a few minutes after the experiment.

L. Stark, P. A. Willis, Gabriella W. Smith

References

1. Dr. David Cogan, Dr. Carl Kupfer, and Dr. Ernst Meyer of Howe Laboratories, Massachusetts Eye and Ear Infirmary, are opthalmologists associated with us in this project.

2. L. Stark and A. A. Sandberg, Model of pupil reflex to light, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M.I.T., January 15, 1963, pp. 237-240.

3. L. R. Young and L. Stark, A sampled-data model for eye-tracking movements, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 370-384.

D. EXPERIMENTS ON DISCRETE CONTROL OF HAND MOVEMENT

Previous reports assembled evidence regarding sampled-data control phenomena in hand motor coordination[1] and also presented a rather complete sampled-data model for eye-movement tracking.[2] Recently, a thesis[3] has been completed which further explores experimental phenomena, which are interpretable in terms of a discrete model.

Unpredictable ramps are used as an input signal, and the rotational load is made

low enough so that the mechanical output elements (muscles, load, apparatus) do not smooth the output signal too drastically. The response of the hand is then a series of steps, as shown in Fig. XVII-9; the hand control system is evidently a discrete position control system. When step amplitudes and times between steps are plotted as a function of ramp velocity as in Fig. XVII-10, it is clear, since only the amplitude and not the sampling period is velocity-dependent, that the nature of the discontinuity is

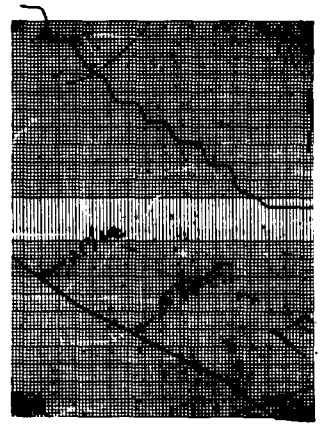


Fig. XVII-9. Random ramp response.

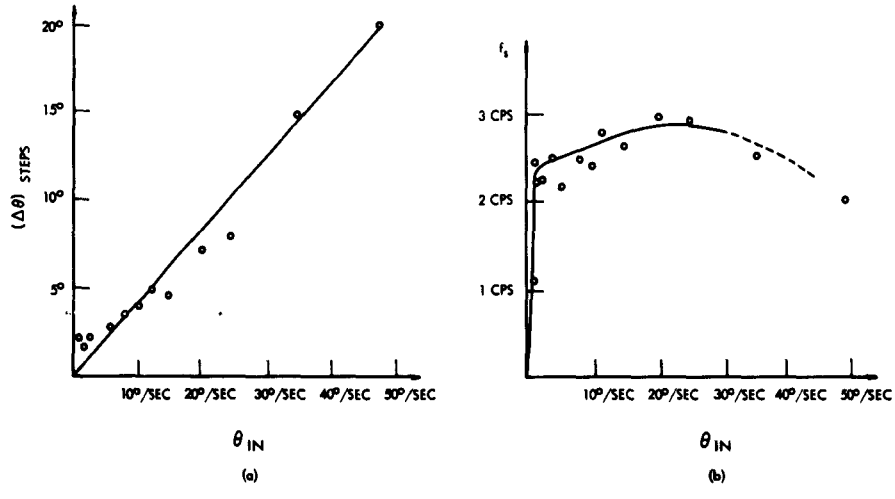


Fig. XVII-10. Dependence of (a) step amplitude, and (b) step frequency as a function of ramp velocity.

not quantization, but rather like a sampled-data system.

Experimentally, varying feedback by means of an environmental clamp is often useful in dissecting a system,[4,5] especially when one can open the loop. When the loop is

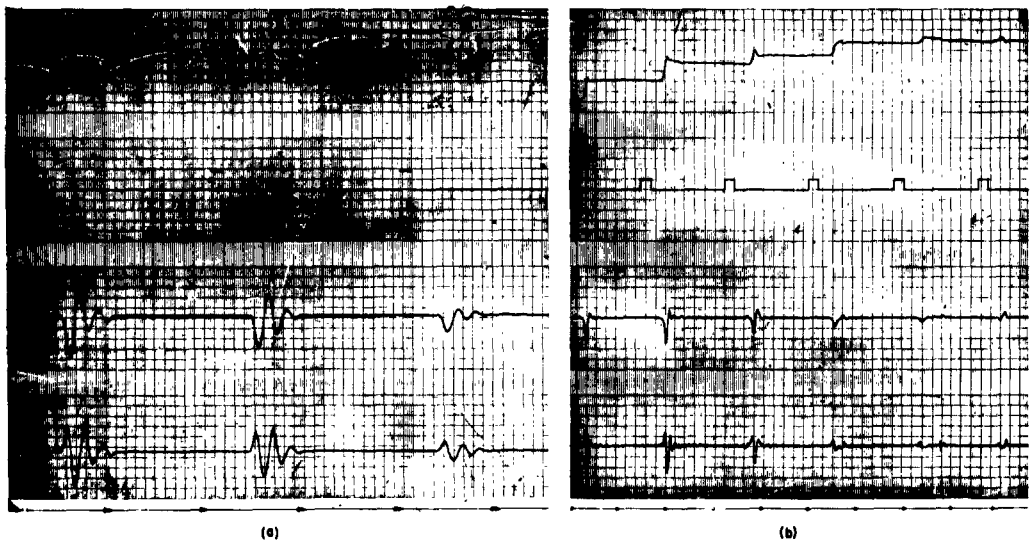


Fig. XVII-11. Pulse response under open-loop conditions.

opened, an unpredictable pulse input should produce a step response. The responses to the initial pulses in experiments such as those displayed in Fig. XVII-11 do indeed approximate steps. Since the inputs are repetitive, and thus eventually predictable, the subject adapts and finally compensates for the change in feedback. The last response to each series of input pulses in Fig. XVII-11 is more nearly a (dynamically limited) pulse. Thus, the position control loop and the adaptive nature of the hand system are both demonstrated in these open-loop experiments.

F. Naves, L. Stark

References

1. L. Stark, Y. Okabe, and P. A. Willis, Sampled-data properties of the human motor coordination system, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 220-223.

2. L. R. Young and L. Stark, A sampled-data model for eye-tracking movements, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 371-384.

3. F. Naves, Sampling or Quantization in the Human Tracking System, S.M. Thesis, Department of Electrical Engineering, M.I.T., January 1963.

4. L. Stark, Environmental clamping of biological systems. The pupil servomechanism, J. Opt. Soc. Am. 52, 925-930 (1962).

5. L. R. Young and L. Stark, Variable feedback experiments supporting a discrete model for eye tracking movements, IRE Transactions on Human Factors (Special Manual Control issue, in press).

XVIII. SHOP NOTES

A. MOCK METRIC IS FUN AND USEFUL TOO

"Four" is a magic number for students, engineers, and scientists in every field in which problems involve comparison or conversion of linear measures expressed in English and metric systems.

There is no obvious compatability between meters equal to 39.37008 inches and inches equal to 0.0254 meter; one familiar with either may find frustration in the other. I have extrapolated that if my English ancestors who made the inch equal to the width of a man's thumb had been foresighted enough to also legislate a man's thumb equal to the fortieth part of a meter, the whole course of history could have been happier!

Nevertheless, we can enjoy a great deal of convenience by imagining the happy ratio We can say with some degree of truth that a meter is forty inches, and we can imagine the forty-inch meter divided and subdivided into the usual metric units. This is to imagine a linear metric system whose units are a bit longer than standard units; a bit so small that for many practical purposes it may be ignored, yet known exactly so that it may be accounted for for other purposes. This is the happy-ratio, first-approximation, pseudometric system that I call "mock metric."

In mock metric, "4" is a magic number:

40 inches = 1 meter	1 inch = $\frac{1}{40}$ meter
4 inches = 1 decimeter	1 inch = $\frac{1}{4}$ decimeter
0.4 inch = 1 centimeter	1 inch = $\frac{10}{4}$ = 2.5 centimeters
0.04 inch = 1 millimeter	1 inch = $\frac{100}{4}$ = 25 millimeters
0.004 inch = 100 microns	1 inch = $\frac{100,000}{4}$ = 25,000 microns.

Any decimal-inch expression can be divided by 4 on mere inspection; the quotient is its mock-metric equivalent in decimeters and may be converted to any other mock-metric unit by shifting the decimal point. For example,

$$7/8 \text{ in.} = 0.875 \text{ in.} = \frac{0.875}{4} = 0.21875 \text{ dm}$$
$$= 2.1875 \text{ cm}$$
$$= 21.875 \text{ mm}$$
$$= 21,875 \ \mu.$$

In mock metric, "3" is a magic number too, but only for dealing with measurements expressed in feet. One need not multiply feet by 12 to obtain inches to be divided by 4;

multiplying feet by 3 gives the same result expressed in mock decimeters.

For example,

$$10 \text{ ft} = 10 \times 3 = 30 \text{ dm}$$
$$= 3 \text{ m}$$

or

$$5 \text{ ft } 9 \text{ in.} = 5.75 \text{ ft} = 5.75 \times 3 = 17.25 \text{ dm}$$
$$= 172.5 \text{ cm}$$

and

$$3 \text{ ft } 4\ 1/2 \text{ in.} = 3 \text{ ft } 4.5 \text{ in.} = 3 \times 3 + \frac{4.5}{4} = 9 + 1.125$$
$$= 10.125 \text{ dm.}$$

These mock-metric expressions so easily obtained are known to be not exact standard metric statements. A feeling for the degree of inexactness is needed for judgment as to whether or not it may be ignored in a given situation. And exact knowledge of the order of the inexactness is needed wherever the inexactness is to be accounted for. Both derive most logically, I think, from considerations of the 1000 standard millimeters that make a standard meter and the 1000 mock millimeters that make a mock meter. Since

$$40 \text{ inches} = 40 \times 25.4 = 1016 \text{ standard millimeters}$$

and

$$40 \text{ inches} = 40 \times 25 = 1000 \text{ mock millimeters,}$$

it follows that

1000 any mock units = 1016 similar standard units.

The order of the inexactness of this first approximation that I call mock metric is 16 parts in 1000 or 1.6 per cent. Obviously, if 1.6 per cent of itself is added to any mock-metric figure, the sum will be the exact standard metric equivalent.

But an easier trick may be to make a second approximation by adding to a mock-metric figure one-sixth of 10 per cent (1/6 · 10 per cent) of itself.

For example,

$$1 \text{ inch} = \frac{1.000}{4} = 0.25 \text{ (mock metric or first approximation)}$$

$$1 \text{ inch} = 0.2500 + \frac{0.025}{6} = 0.25416 \text{ (second approximation).}$$

But we know that

1 inch = 0.25400 standard metric exactly.

Thus it is seen that the inexactness remaining in the second approximation is

reduced to 16 parts in 25,000, which is less than 0.1 per cent and, I think, of approximately the same order as the inexactness of our knowledge of the earth's circumference!

For practical purposes then, in circumstances for which a 1.6 per cent error is not significant or meaningful, the first approximation serves. In more stringent circumstances in which 1.6 per cent may be significant but 0.1 per cent not meaningful, the second approximation serves.

Unusual circumstances that require accuracy better than 1 part in 1000 justify the exact computation instead of the second approximation. And there is no easier way to make the exact computation than to multiply the mock-metric figure by 1.016.

Exact conversion formulas are:

$$\text{meters} = \left(\frac{\text{inches}}{4}\right) 1.016 \text{ standard metric}$$

$$\text{feet} = (3 \times \text{feet})\ 1.016 \text{ standard metric}$$

$$\text{feet and inches} = \left(3 \times \text{feet} + \frac{\text{inches}}{4}\right) 1.016 \text{ standard metric.}$$

The whole philosophy of mock metric is reversible, of course. Any standard metric dimension multiplied by 4 produces a figure that is a first approximation of equivalent inches, or, if divided by 3, a first approximation of equivalent feet, either of which is exactly 101.6 per cent of the exact equivalent. We may divide the first approximation by 1.016 to obtain the exact equivalent, or reduce the first approximation by 1/6 - 10 per cent of itself to obtain the second approximation, with certain knowledge that the remaining error will not be as much as 0.1 per cent.

The decimal point is the same problem in mock metric as in standard metric. In any case, it requires attention to ensure cranking out answers having an error factor of less than 10!

Mock metric defies Parkinson's Law. It is for the man in a hurry, for mental calculation, for solution by inspection, for the fun of empathy with the work you do!

Would you like to see an actual mock-metric scale? Just take any decimal-inch scale and you have it in your hands. Every 0.4-in. interval is a mock centimeter, every 0.040-in. interval a mock millimeter. Only the legend has been arranged to read in terms of inches!

E. C. Ingraham

Author Index